

135042
Shears, Beverly

From: Devi, Sarvamangala
Sent: Tuesday, October 12, 2004 11:10 AM
To: Shears, Beverly
Subject: 10/625,972

Good morning Beverly:

In application 10/625,972, would please performe a sequence search for SEQ ID NO: 4 in commercial and interference databases. Please also run SEQ ID NO: 4 amino acid databases.

Please include an inventors' name search: Phillip I. Tarr; Sima S. Bilge; Thomas E. Besser; James C. Vary

Thanks.

S. DEVI, Ph.D.
AU 1645
Rems - 3C18



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 135042

To: Sarvamangala Devi
Location: REM 3C18
Art Unit: 1645
Friday, October 15, 2004
Case Serial Number: 10/625972

From: Beverly Shears
Location: Remsen Bldg.
RM 1A54
Phone: 571-272-2528
beverly.shears@uspto.gov

Search Notes

Date completed: _____
Searcher: Beverly e 2528
Terminal time: _____
Elapsed time: _____
CPU time: _____
Total time: _____
Number of Searches: _____
Number of Databases: _____

Search Site
____ STIC
____ CM-1
____ Pre-S
Type of Search
____ N.A. Sequence
____ A.A. Sequence
____ Structure
____ Bibliographic

Vendors
____ IG
____ ☒ STN
____ ☒ Dialog
____ APS
____ Geninfo
____ SDC
____ DARC/Questel
____ ☒ Other CGN

This Page Blank (uspto)

GenCore version 5.1.6 -
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: October 13, 2004, 11:44:12 ; Search time 26.5 Seconds
(without alignments)
8147.163 Million cell updates/sec

Title: US-10-625-972-4
Perfect score: 3732
Sequence: 1 atcggaataaccactctggc.....cgctgaactatcagttctga 2091

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO.epool/US10625972/runat_12102004_130106_26853/app.query.fasta_1.2247
-DB=Issued Patents AA -QMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFTW=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10625972@cgn_1_1_25@runat_12102004_130106_26853 -NCPU=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3638	97.5	696	1	US-08-765-081-5
2	3638	97.5	696	3	US-09-098-082-5
3	3638	97.5	696	5	PCT-US95-06994-5
4	3594.5	96.3	703	5	PCT-US95-06994-8
5	3557	95.3	718	5	PCT-US95-06994-7
6	1089.5	29.2	718	5	PCT-US95-06994-6
7	1052.5	28.2	652	1	US-08-765-081-6
8	1052.5	28.2	652	3	US-09-098-082-6
9	1007.5	27.0	677	4	US-09-543-681A-7936
10	956.5	25.6	684	4	US-09-543-681A-4908
11	880.5	23.6	699	4	US-09-489-039A-8133
12	805.5	21.6	703	5	PCT-US95-06994-9

13	800.5	21.4	663	1	US-08-765-081-7	Sequence 7, Appli
14	800.5	21.4	663	3	US-09-098-082-7	Sequence 7, Appli
15	682	18.3	778	4	US-09-489-039A-8455	Sequence 8455, Ap
16	642.5	17.2	759	4	US-09-489-039A-7874	Sequence 7874, Ap
17	641	17.2	795	4	US-09-489-039A-13394	Sequence 13394, A
18	617.5	16.5	761	4	US-09-328-352-5942	Sequence 5942, Ap
19	607.5	16.3	478	4	US-09-252-991A-31866	Sequence 31866, A
20	604.5	16.2	725	4	US-09-668-113A-8	Sequence 8, Appli
21	602	16.1	744	4	US-09-252-991A-29565	Sequence 29565, A
22	537	14.4	651	4	US-09-543-681A-8284	Sequence 8284, Ap
23	400	10.7	617	4	US-09-252-991A-30921	Sequence 30921, A
24	396.5	10.6	1086	4	US-09-252-991A-25051	Sequence 25051, A
25	383.5	10.3	643	4	US-09-328-352-5146	Sequence 5146, Ap
26	371.5	10.0	498	4	US-09-489-039A-12187	Sequence 12187, A
27	360.5	9.7	478	4	US-09-252-991A-31603	Sequence 31603, A
28	359.5	9.6	789	4	US-09-252-991A-27011	Sequence 27011, A
29	349	9.4	704	4	US-09-489-039A-9841	Sequence 9841, Ap
30	340	9.1	704	4	US-09-328-352-5537	Sequence 5537, Ap
31	339.5	9.1	695	4	US-09-489-039A-14338	Sequence 14338, A
32	316.5	8.5	803	4	US-09-252-991A-28469	Sequence 28469, A
33	315	8.4	685	4	US-09-543-681A-4482	Sequence 4482, Ap
34	315	8.4	743	4	US-09-328-352-5073	Sequence 5073, Ap
35	312.5	8.4	790	3	US-08-817-707-6	Sequence 6, Appli
36	311.5	8.3	718	4	US-09-252-991A-29432	Sequence 29432, A
37	311	8.3	844	4	US-09-252-991A-25037	Sequence 25037, A
38	307.5	8.2	727	4	US-09-489-039A-14168	Sequence 14168, A
39	304	8.1	791	3	US-08-537-361E-4	Sequence 4, Appli
40	303	8.1	903	4	US-09-252-991A-24977	Sequence 24977, A
41	301	8.1	790	3	US-08-817-707-8	Sequence 8, Appli
42	301	8.1	791	3	US-08-817-707-4	Sequence 4, Appli
43	298.5	8.0	792	3	US-08-537-361E-2	Sequence 2, Appli
44	294.5	7.9	766	4	US-09-328-352-8230	Sequence 8230, Ap
45	293.5	7.9	792	1	US-08-326-670A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-08-765-081-5
Sequence 5, Application US/08765081
Patent No. 5798260
GENERAL INFORMATION:
APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
STREET: 2800 Pacific First Center, 1420 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
COMPUTER: IBM PC compatible/Pentium
OPERATING SYSTEM: MS-Windows 3.1
SOFTWARE: Word for Windows-6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/765,081
FILING DATE: March 26, 1997
PRIOR APPLICATION DATA: PCT/US95/06994
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995
APPLICATION NUMBER: US 08/265,714
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997
REFERENCE/DOCKET NUMBER: CHOR-1-10286
TELECOMMUNICATION INFORMATION: 1-206-682-8100; 1-206-224-0718 (direct)
TELEPHONE: 1-206-682-8100; 1-206-224-0779
TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
 LENGTH: 696 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-765-081-5

Alignment Scores:
 Pred. No.: 0 Length: 696
 Score: 3638.00 Matches: 696
 Percent Similarity: 100.00% Conservatives: 0
 Best local Similarity: 100.00% Mismatches: 0
 Query Match: 97.48% Indels: 0
 DB: 1 Gaps: 0

US-10-625-972-4 (1-2091) x US-08-765-081-5 (1-696)

QY	1	ATCGGAATTAACCACTCTGGCTTCGGTAGTCATTCCTGTCGGATTTTCAGCAGCAGC	60
DB	1	MetArgIleThrThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSerSer	20
QY	61	ATAGCTGCTGCAGAGGATGTGATGTTGTCCTGGCATCCGGCTATGAGAAAAGCTGACT	120
DB	21	IleAlaAlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLysLeuThr	40
QY	121	AACGAGCCGCGAGTGTCTGTGATAGCCAGGAGGAATTCAGTCCAGCCAGTACCAC	180
DB	41	AsnAlaAlaAlaSerValSerValIleSerGlnGluLeuGlnSerSerGlnTyrHis	60
QY	181	GATCTCGCGGAGCTCTGAGATCAGTAGAGGTGGGATGTTGAAAGTGGTACGGTAAA	240
DB	61	AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys	80
QY	241	ACCGAGGCTCGAATACAGATCCGAGGAATGCCAGCCAGTTACACCTGATCTGATT	300
DB	81	ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyrThrLeuIleLeuIle	100
QY	301	GATGTGTTGTCAGGCGGAGCAGTACGCTGCTCCACCGTTCCTCCAGCTGATGAT	360
DB	101	AspGlyValArgGlnGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn	120
QY	361	ACCGGTTTCATCCCTCTGCGCCCATTCAGCGTATTGAGTTATCAGGGGCGCGATG	420
DB	121	ThrGlyPheMetProProLeuAlaAlaIleGluArgIleGluValIleArgGlyProMet	140
QY	421	TCGACACTGTATGCTCTGATCGATGGCGGTGGTGAATATCATTCACGAGAAGAT	480
DB	141	SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAsnIleIleThrArgLysAsn	160
QY	481	GCAGACAAATGGCTCTCTCCGTCAATCAGGGCTGAAATCTGCAGGAAAGCAACAATGG	540
DB	161	AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnSerAsnLysTrp	180
QY	541	GGTACACAGCCAGTTTAATTTCTGGAGCAGTGCTCCCTTGTGGATGATCTGTGAGC	600
DB	181	GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspSerValSer	200
QY	601	CTGACAGTACGGGTAGCACACAACAGCGTCAGGGTTCATCGGTACATCACTCAGCGAT	660
DB	201	LeuGlnValArgLysThrGlnGlnArgGlnGlySerSerValThrSerLeuSerAsp	220
QY	661	ACAGCAGCAGCGGTATTCCTTATCCACGAGTACACAGAATTATAATCTTGTGGCAGT	720
DB	221	ThrAlaGlyThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg	240
QY	721	CTTGACTGAGGCTCGGACAGATGTCTGTTGTATGATATGATACACCCCGCAG	780
DB	241	LeuAspTrpLysAlaSerGlnAspValLeuTrpPheAsnMetAspThrThrArgGln	260
QY	781	CGTTATGATAACCGGATCGGCACTGGGGAGTCTGACGGGGGATATGACCGGACCTG	840
DB	261	ArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGlyTyrAspArgThrLeu	280

QY	841	CGCTATGAGCGAAACAAAAATTTTCAGTGGCTATGATCATCTTTCACCTTCGGAACATGG	900
DB	281	ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspHisThrPheThrPheGlyThrTrp	300
QY	901	AAATCGTATCTGAACCTGGAAACGAGACAGAAAAATAAGGTGCTGAGCTTGTACCATGTA	960
DB	301	LysSerTyrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal	320
QY	961	CTGAAGCGCACAAATGGGGCTTCGGTTCAGCCCGGGAGCTTAAGAAATCGAACCTT	1020
DB	321	LeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgLysLysLysGluSerAsnLeu	340
QY	1021	ATCCTGAATTCATTACTGCTTACCCCTCTGGGAGAAATCTCATCTGCTTACGTTGGGGG	1080
DB	341	IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly	360
QY	1081	GAGTTTCAGAGCTGCCCATGAAGACGAGTGTCTTCCAGACACAGTGAACATTTTC	1140
DB	361	GluPheGlnSerSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe	380
QY	1141	CGCAGAAAAAGCTGGTGGTATTTGCTGAGGATGAGTGGCATCTCACGGATGCACTTGG	1200
DB	381	ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspAlaLeuAla	400
QY	1201	CTGACTCGCGGACGCGCTATGAACATCATGAGCAATTCGGGGGACACTTCAGTCGGCT	1260
DB	401	LeuThrAlaGlySerArgTyrGluHisHisGluGlnPheGlyGlyHisPheSerProArg	420
QY	1261	GCATATCTGCTCGGATGTCGAGATGCTGACCGCTCAAAAGGCGGTGTGACCCAGGA	1320
DB	421	AlaTyrLeuValTrpAspValAlaAspAlaTrpThrLeuLysGlyGlyValThrGly	440
QY	1321	TATAAGCCACCCAGAAATGGGAGCTACATAAAGGATAGTGGTGTGTCCGGCAGGA	1380
DB	441	TyrLysAlaProArgMetGlyGlnLeuHisLysGlyIleSerGlyValSerGlyGlnGly	460
QY	1381	AAACAAATCTACTGTTAAACCCCGACCTGAAGCGGAGAGAGAGCGTCAAGTATCAGGCT	1440
DB	461	LysThrAsnLeuLeuGlyAsnProAspLeuLysProGluLysValSerTyrGluAla	480
QY	1441	GGGCTTATTACGATAACCCCGCTCTGAATGCCAATGTCCACAGGTTTATGACTGAC	1500
DB	481	GlyValTyrTyrAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp	500
QY	1501	TTCTCCAAAGAGTGTCTCTTATTCATTAATGATACACCAATAGCTATGTAACAGC	1560
DB	501	PheSerAsnLysIleValSerTyrSerIleAsnAspAsnThrAsnSerTyrValAsnSer	520
QY	1561	GGAAAGCGCGGTGTCACCGTGTGAAATTTGCCGCGCACATTCGCCCTGTGGTCAGAGAT	1620
DB	521	GlyLysAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp	540
QY	1621	GTACCGCTGTACTGAAATTTACACCTGGACCCGAGTGAACACCAAGTATGCTGATACAAA	1680
DB	541	ValThrLeuSerLeuAsnTyrThrTrpThrArgSerGluGlnArgAspGlyAspAsnLys	560
QY	1681	GGTGGCCCTCAGTTATACCCCTGAAACACATGGTGAATGCGAACTGAACCTGGCAGATC	1740
DB	561	GlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle	580
QY	1741	ACGGAAGAGTGGCATCATGCTGGTGGTGGCTTATTCGGCGGAAACACACGTTTACC	1800
DB	581	ThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgLysThrProArgPheThr	600
QY	1801	CAGAATTATTCCTCAGCTGAGCGCTGACAGAGAAAGTGTATGATGAGAAAGGACAATAC	1860
DB	601	GlnAsnTyrSerSerLeuSerAlaValGlnLysLysValTyrAspGluLysGlyGluTyr	620
QY	1861	CTCAAAGCTGACCGGTGGTGGATCGAGTCTGTCTGTGAAAGATGACCGATGCCCTGACG	1920
DB	621	LeuLysAlaTrpThrValValAspAlaGlyLeuSerTrpLysMetThrAspAlaLeuThr	640
QY	1921	CTGAATGCTCGCGTGAATAACCTGCTCAACAGGATTACAGTGCCTGACGCTGTACAGT	1980

Db 641 LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyrSerAspValSerLeuTyrSer 560
Qy 1381 GCCGTAAGAGTACGCTGTATGCGGTGATTACTTCCAGACGGGATCATCAACACAGCA 2040
Db 561 AlaglyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThrThrGly 580
Qy 2041 TATGTGATCTAGCGAAATTAATCTGATGTCCTGACATCATCAGTTC 2088
Db 681 TyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe 696
RESULT 2
US-09-098-082-5
; Sequence 5, Application US/09098082
; Patent No. 6040421
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible/Pentium II
; OPERATING SYSTEM: MS-Windows 95
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,082
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/765,081
; FILING DATE: March 26, 1997
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: CHOR-1-12402
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 696 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-098-082-5
Alignment Scores:
Pred. No.: 0 Length: 696
Score: 3638.00 Matches: 696
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 97.48% Indels: 0
DB: 3 Gaps: 0
US-10-625-972-4 (1-2091) x US-09-098-082-5 (1-696)
Qy 1 ATGCGAATAACCACTGCTGCTCCGTAGTCATTCCTGCTCGGATTTTCAGCCAGCAGC 60
Db 1 MetA-gileThrThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSerSer 20
Qy 61 ATAGCTGCTGACAGAGATGATGATGTCCTCGGATCCGGGTATGAGAAAGCTGACT 120
Db 21 IleAlaAlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLysLeuThr 40

Qy 121 AACGACGCGCGAGTGTTCCTGTGATTAGCAGAGGAATTCAGTCCAGCCAGTACCAC 180
Db 41 AsnAlaAlaAlaSerValSerValIleSerGlnGluGluLeuGlnSerSerGlnTyrHis 60
Qy 181 GATCTGCGCGAGGCTCTCAGATCAGTAGAGGTGTGGATGTGAAAGTGGTACGGGTAAA 240
Db 61 AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys 80
Qy 241 ACCGAGGCGCTGGAATCAGATCCGAGGAATGCCAGCCAGTTACACCTCATATCTGATT 300
Db 81 ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyrThrLeuIleLeuIle 100
Qy 301 GATGTGTTCGTACGAGCGGAGACAGTACGCTGACTCCCAACGGTTCCTCCCATGAAT 360
Db 101 AspGlyValArgGlnGlyGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
Qy 361 ACCGGTTCATCCCTCTGCTGCGCCCATTTGAGCGTATTGAGTTATCAGGGGCGCGATG 420
Db 121 ThrGlyPheMetProProLeuAlaIleGluArgIleGluValIleArgGlyProMet 140
Qy 421 TCACACTGTATGGCTCTGATCGGATGGCGGTGTGTGAATATCATCACCAGAAAGAT 480
Db 141 SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAsnIleIleThrArgLysAsn 160
Qy 481 GCAGACAAATGCTCTCTTCCTCAATGAGGGCTGAATCTCCAGAAAGCAACAAATCG 540
Db 161 AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp 180
Qy 541 GGTACAGCAGCAGTAAATTTCTGGAGCAGTGTCCCTTGTGGATGATTCTCTGACG 600
Db 181 GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspSerValSer 200
Qy 601 CTCAGGTACGCGGTAGCACACAGCCTCAGGTTTCATCGGTACATCCTCAGCCGAT 560
Db 201 LeuGlnValArgGlySerThrGlnGlnArgGlnGlySerSerValThrSerLeuSerAsp 220
Qy 661 ACAGCAGCGCAGCGTATTCCTTATCCACGGAGTACAGAAATTAATCTTCGTGACGT 720
Db 221 ThrAlaGlyThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg 240
Qy 721 CTGACTGGAAGCGCTCGGAGCAGGATGTCTCTGTTGTATGATGATACCCCGGCGAG 780
Db 241 LeuAspTrpLysAlaSerGluGlnAspValLeuTrpPheAspMetAspThrThrArgGln 260
Qy 781 CGTTATGATAACCGGATCGGCACTGGGAGTCTGACGGGGGATATCAGCGGACCCG 840
Db 261 ArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGlyTyrAspArgThrLeu 280
Qy 841 CGCTATGAGCAACAAATTTTCAGCTGCTGCTATGATCATACTTTCACCTTCGGAACATGG 900
Db 281 ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspHisThrPheThrPheGlyThrTrp 300
Qy 901 AAATCGTATCTGAATCTGAACTGGAACGAGACAGAAATAAAGGTCTGTGAGTGTAC 960
Db 301 LysSerTyrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal 320
Qy 961 CTGAACGCGCACAATGGGGCTTCGCGTTCAGCCCGGAGGCTTAAGGAATCGAACCTT 1020
Db 321 LeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu 340
Qy 1021 ATCTGAAATTCATTACTGCTTACCCTCTGGGAGAATCTCATCTGTTACGTGGGGGGC 1080
Db 341 IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly 360
Qy 1081 GAGTTTCAGAGCTCGTCCATGAAGACGAGGATTTGCTCTCCAGCAGCAGGTGAACCTTC 1140
Db 361 GluPheGlnSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe 380
Qy 1141 CGGCAGAAAAGCTGCTCGGTATTTGCTGAGGATCAGTGGCATCTCAGGGATGCACTGCG 1200
Db 381 ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspAlaLeuAla 400

1201 CTGACTCGGGCAGCCGCTATGACATCATGACCAATTCGGGGACACTTCAGTCGGCT 1260
 401 LeuThrAlaGlySerArgTyrGluHisHisGluGlnPheGlyHisPheSerProArg 420
 1261 GCATATCTGTCTGGGATGGCAGATGCTGACCGCTGAAAGCGGTGTGACCAACGGGA 1320
 421 AlaTyrLeuValTrpAspValAlaAspAlaTrpThrLeuLysGlyGlyValThrThrGly 440
 1321 TATAAGCCACCAAGATGGGCGAGCTACATAAAGGATAGTGGTGTCGGGCGAGGGA 1380
 441 TyrLysAlaProArgMetGlyGlnLeuHisLysGlyLysGlyValSerGlyGlnGly 460
 1381 AAAACAAATCTACTTGGTAAACCCGACCTGAAAGCCGAGAGCGTCAAGTTATGAGCT 1440
 461 LysThrAsnLeuLeuGlyAsnProAspLeuLysProGluGluSerValSerTyrGluAla 480
 1441 GGGGTGTATTACGATAACCCCGCGTCTGAATGCCAATGTACAGGTTTTATGACTGAC 1500
 481 GlyValTyrTyrAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp 500
 1501 TTCTCCAAACAGATGCTCTTATTCATTAATGATACCAATAGCTATGTAACAGC 1560
 501 PheSerAsnLysLysValSerTyrSerIleAsnAspAsnThrAsnSerTyrValAsnSer 520
 1561 GGAAGCCCGGTTGCACGGTGTGAATTTGCCGCGACATTCGCCGCTGTGTGTCAGAGAT 1620
 521 GlyLysAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp 540
 1621 CTCACGCTCTACTGATTAACCTGGACCGCCGAAGTGAACACAGTGGTGATTAACAAA 1680
 541 ValThrLeuSerLeuAsnTyrThrTrpThrArgSerGluGlnArgAspGlyAspAsnLys 560
 1681 GGTGCCCGCTCAGTTATATACCCCTGAACACATGTTGAATGCGAACTGAACCTGGCAGATC 1740
 561 GlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle 580
 1741 ACCGAAGCTGGCATCATGGTGGTGGTCCGCTTATCGGGGAAACACCGCTTTCACC 1800
 581 ThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgGlyLysThrProArgPheThr 600
 1801 CAGAAATTATCTGCTACTGAGCGCTGTACAGAGAAGTGTATGATGAGAAGAGGACATAC 1860
 601 GlnAsnTyrSerSerLeuSerAlaValGlnLysLysValTyrAspGluLysGlyLysThr 620
 1861 CTGAAAGCTGACCGTGGTGGATCAGCTGTCTGCGTGAAGATGACCGATCCCGTACG 1920
 621 LeuLysAlaTrpThrValValAspAlaGlyLeuSerTrpLysMetThrAspAlaLeuThr 640
 1921 CTGAATGCTGGCGTGAATTAACCTGTCAACAGGATTAACAGTGGAGCTGTACAGT 1980
 641 LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyrSerAspValSerLeuTyrSer 660
 1981 GCCGTGAAGATACGCTGTATGCCGCTGATTACTTCCAGACGGGATCATCAACACAGGA 2040
 661 AlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThrThrGly 680
 2041 TATGTGATACCTGACGGAATTAACGATGTCGCTGGAATCATCACTCATCAGTTC 2088
 681 TyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe 696

RESULT 3
 PCT-US95-06994-5
 ; Sequence 5, Application PC/TUS9506994
 ; GENERAL INFORMATION:
 ; APPLICANT: Children's Hospital & Medical Center
 ; APPLICANT: University of Washington
 ; APPLICANT: Washington State University Research Foundation
 ; APPLICANT: TARR, PHILLIP I
 ; APPLICANT: BILGE, SIMA S
 ; APPLICANT: BESSER, THOMAS E
 ; APPLICANT: VARY JR, JAMES C
 ; TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
 ; NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:
 ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
 STREET: SUITE 2800, 1420 FIFTH AVENUE
 CITY: SEATTLE
 STATE: WASHINGTON
 COUNTRY: USA
 ZIP: WA 98101
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/06994
 FILING DATE: 07-JUN-95
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/265,714
 FILING DATE: 24-JUN-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: BRODERICK, THOMAS F
 REGISTRATION NUMBER: 31,332
 REFERENCE/DOCKET NUMBER: CHOR-18591
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 682 8100
 TELEFAX: (206) 224 0779
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 696 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-06994-5

Alignment Scores:
 Pred. No.: 0 Length: 696
 Score: 3638.00 Matches: 696
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 97.48% Indels: 0
 DB: 5 Gaps: 0

US-10-625-972-4 (1-2091) x PCT-US95-06994-5 (1-696)

QY 1 ATCGAATAACCACTCTGGCTTCGCTAGTCAATTCCTGTCGGATTTTCAGCCAGCAGC 60
 Db 1 MetArgIleThrThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSerSer 20
 QY 61 ATAGCTGCTGCAGAGCATGTGATGTTCTCGGCATCCGCTATGAGAAAGCTGACT 120
 Db 21 IleAlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLeuThr 40
 QY 121 AACGACCCGCGCAGTGTCTGTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCAC 180
 Db 41 AsnAlaAlaSerValSerValIleSerGlnGluLeuGlnSerSerGlnTyrHis 60
 QY 181 GATCTGGCGGAGCTCTGAGATCAGTAGGGGTGTGGATGTTGAAGTGTTCGGGTAAA 240
 Db 61 AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys 80
 QY 241 ACCGGAGGGCTGAAATCAGCATCCGAGGAATGCCAGCAGTTCACGCTGATCTGATT 300
 Db 81 ThrGlyLysLeuGluIleSerIleArgGlyMetProAlaSerTyrThrIleIleLeu 100
 QY 301 GATGTTGTTCTCAGCGCGGAAGCAGTGCAGTGCATCCCAACGGTTTTTCTGCCATGAAT 360
 Db 101 AspGlyValArgGlnGlyGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
 QY 361 ACCGGTTTCATGCCCTCTGCGCCCATGAGCGTATTGAGTTATCAGGGGCGCGATG 420
 Db 121 ThrGlyPheMetProLeuAlaAlaIleGluArgIleGluValIleArgGlyProMet 140
 QY 421 TCCACATGTATGGCTCTGATGCGATGGCGGTGTGGTGAATATCATTTACCAGAAAGAAT 480

141 SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAsnLeuLeuThrArgLysAsn 160
481 GCAGACAAATGGCTCTTCCGTCATCGCAGGGCTGAATCTGCAGGAAAGCAAAATG 540
161 AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp 180
541 GGTAAACAGCAGCCAGTTAATCTCTGAGAGAGTGGTCCCTTGTGATGATTTCTGCAGC 600
181 GlyAsnSerSerGlnPheAsnProTrpSerSerGlyProLeuValAspSerValSer 200
601 CTGCAGGTAGCGGTAGCAGACACACAGCGTCAGGGTTCATCGCTCACATCACCTGAGCGAT 660
201 LeuGlnValArgGlySerThrGlnGlnArgGlnGlySerSerValThrSerLeuSerAsp 220
661 ACAGCAGCGCGGTATTCCTTATCCACCGAGTTCAGAAATATATATCTTGTGTGACGT 720
221 ThrAlaGlyThrArgLysProTrpProTrpGluSerGlnAsnTyrAsnLeuGlyAlaArg 240
721 CTTCAGCTGGAAGCGTGGAGCAGGATGCTCTGTTTGCATATGATGATACACCGGCAG 780
241 LeuAspTrpLysAlaSerGluGlnAspValLeuTrpPheAspMetAspThrThrArgGln 260
781 CTTATGATATACCGGGATGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCGCTG 840
261 ArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGlyTyrAspArgThrLeu 280
841 CGCTATGAGCGAAACAAATTTACGCTGGCTATCATCTACTTTACCTTCGGACATGG 900
281 ArgTyrGluArgAsnLysLysSerAlaGlyTyrAspHisThrPheThrPheGlyThrTrp 300
901 AAATCGTATCTGAATCGAAGCAGACAGAAATAAAGTCTGAGCTTGACGAGTGTA 960
301 LysSerTyrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal 320
961 CTGAAGCCGCAAAATGGGGCTGCGCGTCCAGCGCGGAGCTTAGGAATCGAACTT 1020
321 LeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu 340
1021 ATCTGATTTCTATCTCTACCTCTGGAGAACTCTCATCTGTTTACGTCGGGGGC 1080
341 IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly 360
1081 GAGTTTCAGAGCTCTCCATGAAGACGAGGTGTCCTTGCACGACAGGTGAACCTTTC 1140
361 GluPheGlnSerSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe 380
1141 CGGCAAGAAAGCTGCTCGTATTTCTGAGATGAGTGGCTCTCACGGATGCACTTGGC 1200
381 ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspAlaLeuAla 400
1201 CTGACTGGGGCAGCGCTATGAACATCATGACCAATTCGGGGACACTTCAGTCCGCT 1260
401 LeuThrAlaGlySerArgTyrGluHisGluGlnPheGlyGlyHisPheSerProArg 420
1261 GCATATCTGGTGGGATGTGCAGATGCTGGAGCTGGAAGCGGTGTGACCCGGA 1320
421 AlaTyrLeuValTrpAspValAlaAspAlaTrpThrLeuLysGlyGlyValThrThrGly 440
1321 TATAGGCAACCCAGATGGGCGAGCTACATAAGGATTTAGTGTGTCTCGGGCAGGGA 1380
441 TyrLysAlaProArgMetGlyGlnLeuHisLysGlyLysSerGlyValSerGlyGlnGly 460
1381 AAAACAAATCTACTTGGTAAACCCGACCTGAAGCCGGAAGAGAGCGTCACTTATGAGCT 1440
461 LysThrAsnLeuLeuGlyAsnProAspLeuLysProGluGluSerValSerTyrGluAla 480
1441 GGGGTGTATACGATTAACCCCGCGTCTGAATGCCAATGTACAGGTTTATGACTGAC 1500
481 GlyValTyrAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp 500
1501 TTCTCCAAACAGATTCTCTTATTCATTAATGATAAATGATAACCAATAGCTATGATAACAGC 1560

501 PheSerAsnLysIleValSerTyrSerIleAsnAspAsnThrAsnSerTyrValAsnSer 520
1561 GGAAAGGCGCGGTTCACGGTGTGGAATTCGCGGCACATTGCGCTGTGTCTCAGAGGAT 1620
521 GlyLysAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp 540
1621 GTACAGCTGTCTCACTGAATTCACCTGACCCGCAAGTGAACACAGTGTGTTGATTAACAAA 1680
541 ValThrLeuSerLeuAsnTyrThrTrpThrArgSerGluGlnArgAspGlyAspAsnLys 560
1681 GTGCGCGCTGAGTTATACCCCTGAACACATGTTGTAATGCGAACTGAACCTGCGCAGATC 1740
561 GlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle 580
1741 ACCGAAGAGTGGCATCATGCTGGTGGTGGTTCGCGGAAACACCACTGTTTACC 1800
581 ThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgGlyLysThrProArgPheThr 600
1801 CAGAAATATCTCTCATCTGAGCGCTGTACAGAAGAAGTGTATGATGAGAAGAGCAATAC 1860
601 GlnAsnTyrSerSerLeuSerAlaValGlnLysLysValTyrAspGluLysGlyLysTrp 620
1861 CTGAAGCTGACGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1920
621 LeuLysAlaTrpThrValValAspAlaGlyLeuSerTrpLysMetThrAspAlaLeuThr 640
1921 CTGAATGCTGCGTGAATACTGCTCAACAGAGTATACAGTCACTGAGCTGAGCTGACAGT 1980
641 LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyrSerAspValSerLeuTyrSer 660
1981 GCGGTAAAGAGTACGCTGTATCGCGTGAATTAATTCACAGCGGATCATCAACACAGGA 2040
661 AlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThrThrGly 680
2041 TATGTGATACCTGAGCAAAATTAAGTGTGCTGCTGCACTGCACTGCACTGCACTGCACT 2088
681 TyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe 696

RESULT 4
PCT-US95-06994-8
Sequence 8, Application PC/TUS9506994
GENERAL INFORMATION:
APPLICANT: Children's Hospital & Medical Center
APPLICANT: University of Washington
APPLICANT: Washington State University Research Foundation
APPLICANT: TARR, PHILLIP I
APPLICANT: BILGE, SIMA S
APPLICANT: BESSER, THOMAS E
APPLICANT: VARY JR, JAMES C
TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
STREET: SUITE 2800, 1420 FIFTH AVENUE
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: WA 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: 07-JUN-95
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,714
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BRODERICK, THOMAS F
REGISTRATION NUMBER: 31,332

REFERENCE/DOCKET NUMBER: CHOR-18591

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 682 8100
 TELEFAX: (206) 224 0779
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 703 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 DESCRIPTION: E. coli O157:H7 adhesin amino acid sequence,
 wherein "Xaa" residues represent gaps
 DESCRIPTION: introduced to facilitate best alignment with
 DESCRIPTION: SEQ ID NO:9.
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Escherichia coli O157:H7
 STRAIN: 86-24 NALR
 PCT-US95-06994-8

Alignment Scores:

Pred. No.: 0 Length: 703
 Score: 394.50 Matches: 695
 Percent Similarity: 98.86% Conservative: 0
 Best Local Similarity: 98.86% Mismatches: 1
 Query Match: 96.32% Indels: 7
 DB: 5 Gaps: 3

US-10-625-972-4 (1-2091) x PCT-US95-06994-8 (1-703)

QY	1	ATCGGAATACCACTCTGCTCCGTCAGTCATCTCCGTCGAGTTTCACGACGAGC	60
Db	1	MetArgileThrThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSerSer	20
QY	61	ATA-----CCTGCTCAGAGGATGATGATGCTCGGCAATCCGGCTAT	105
Db	21	Ile*****AlaAlaAlaGluAspValMetIleValSerAlaSerGlyTyr	40
QY	106	GAGAAAGCTGACTAACCGCCGCTGCTGTCGATTCAGGAGGAAATTCGAG	165
Db	41	GluLysLysLeuThrAsnAlaAlaAlaSerValSerValIleSerGlnGluLeuGln	60
QY	166	TCAGCAGTACACGATCTGCGAGGCTCTGACATCAGTAGAGGCTGCGATTCGAA	225
Db	61	SerSerGlnTyrHisAspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGlu	80
QY	226	ACTGTACGGGTAAACCGGGGTGGAATACAGATCCGAGGAATCCAGCCAGTTAC	285
Db	81	SerGlyThrGlyLysThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyr	100
QY	286	ACGCTGATCTGATGATGGTGTCTGTCAGGCGGAGACAGTACGTCGACCTCCACGGT	345
Db	101	ThrLeuIleLeuIleAspGlyValArgGlnGlySerSerAspValThrProAsnGly	120
QY	346	TTTTCTGTCATGAATACCGGGTTCATGCCCTCTGGCCGCTTACGCGTATTCAGGCTT	405
Db	121	PheSerAlaMetAsnThrGlyPheMetProLeuAlaAlaIleGluArgIleGluVal	140
QY	406	ATCAGGGGCGGATGTCACACTGTATGCTCTGATCGATGCGGCGGTGGTGAATATC	465
Db	141	IleArgGlyProMetSerThrLeuTyrGlySerAspAlaMetGlyValValAsnIle	160
QY	466	ATTACAGAAAGATGCAGACAAATGGCTCTCTCCGTCATGACGAGGCTGAATCTCAG	525
Db	161	IleThrArgLysAsnAlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGln	180
QY	526	GAAAGCAAAATGGGTAAACAGACGAGCTTAAATTCGGAGCAGTGGTCCCTTTGTG	585
Db	181	GluSerAsnLysTrpGlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuVal	200
QY	586	GATGATCTGTCAGCTCAGGTACGGGTAGCACACACAGCTCAGGCTTCATCGGTC	645
Db	201	AspAspSerValSerLeuGlnValArgGlySerThrGlnArgGlnGlySerSerVal	220

QY	646	ACATCACTAGCGGATACAGACGAGCACCGGTATTCCTTATCCACGGAGTCCACAGAATTAT	705
Db	221	ThrSerLeuSerAspThrAlaGlyThrArgIleProTyrProThrGluSerGlnAsnTyr	240
QY	706	AATCTTCGTGCAGCTCTTGACTGGAGCGCTCGGACGAGGATGTCCTCTGTTGATATG	765
Db	241	AsnLeuGlyAlaArgLeuAspTrpLysAlaSerGlnGlnAspValLeuTrpPheAspMet	260
QY	766	GATACACCCCGCAGCGCTTATGATAACCGGATGCGCACTGGGGAGTCTGACGGGGGA	825
Db	261	AspThrThrArgGlnArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGly	280
QY	826	TATGACCGGACCTCGCTATGACGCAACAAATTCAGCTGGCTATGATCATCTTTC	885
Db	281	TyrAspArgThrLeuArgTyrGluArgAsnLysIleSerAlaGlyTyrAspGlyThrPhe	300
QY	886	ACCTTCGGAACATCGAAATCGTATCTGAACGAGACGACAGAAAAATAAGGTCGTGAG	945
Db	301	ThrPheGlyThrTrpLysSerTyrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGlu	320
QY	946	CTTGACGCGAGTGTACTGAAGCGCGCAAAATGGGGCTTCGCGTCAGCCGCGGAGCTT	1005
Db	321	LeuValArgSerValLeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeu	340
QY	1006	AAGGAATCGAACCTTATCTGAATTCATTACTGCTTACCCCTCTGGGAGAAATCTCATCG	1065
Db	341	LysGluSerAsnLeuIleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeu	360
QY	1066	GTTACGCTGGGGCGGAGTTTCAGACTCCTCATGAAGACGAGTTCCTCTTCCACG	1125
Db	361	ValThrValGlyGlyGluPheGlnSerSerMetLysAspGlyValValLeuAlaSer	380
QY	1126	ACAGGTGAACTTTCGCGCAGCAAAAGC---TGGTCGTATTTGCTCAGAGATGAGTGGAT	1182
Db	381	ThrGlyGluThrPheArgGlnLysSer***TrpSerValPheAlaGluAspGluTrpHis	400
QY	1183	CTCACGATGCATCTCGCTGCTGCGGCGCAGCTATGAAACATCATGAGCAATTCGGG	1242
Db	401	LeuThrAspAlaLeuAlaLeuThrAlaGlySerArgTyrGluHisGluGlnPheGly	420
QY	1243	GGACACTCAGTCCGCTGCATATCTGGTCTGGATGTGGCAGATGCCCTGGACGCTGAA	1302
Db	421	GlyHisPheSerProArgAlaTyrLeuValTrpAspValAlaAspAlaTrpThrLeuLys	440
QY	1303	GGCGGTGTACACACGCGGATATAAGGCACCCAGAAATGGGCGAGCTACATAAGGGATTAGT	1362
Db	441	GlyGlyValThrThrGlyTyrLysAlaProArgMetGlyGlnLeuHisLysGlyIleSer	460
QY	1363	GGTGTCTCCGGCAGGAGAAAAACAATCTACTTGGTAACCCCGACTGAAGCCGAGAG	1422
Db	461	GlyValSerGlyGlnGlyLysThrAsnLeuLeuGlyAsnProAspLeuLysProGluGlu	480
QY	1423	AGCGTCAGTTATGAGCTGGGTGATTACGATAACCCCGCGGCTCGAATCCCAATGTC	1482
Db	481	SerValSerTyrGluAlaGlyValTyrTyrAspAsnProAlaGlyLeuAsnAlaVal	500
QY	1483	ACAGGTTTTATGACTGACTTCTCCAAACAGATT---GTCTCTTATTCCATAAATGATAAC	1539
Db	501	ThrGlyPheMetThrAspPheSerAsnLysIle***ValSerTyrSerIleAsnAspAsn	520
QY	1540	ACCAATAGCTATGTAACAGCGGAAGCCCGGTTCACGGTGTGGATTTGCCGGCACA	1599
Db	521	ThrAsnSerTyrValAsnSerGlyLysAlaArgLeuHisGlyValGluPheAlaGlyThr	540
QY	1600	TTGCCGCTGTGTGTCAGAGGATGTCAGCTGTCACCTGCTCACTGAATTACACCTGACCCGAGTCAA	1659
Db	541	LeuProLeuTrpSerGluAspValThrLeuSerLeuAsnTyrThrTrpThrArgSerGlu	560
QY	1660	CAACCTGATGCTGATAACAAAGGTCCGCGTGTGATTTATACCTCTGACACATGCTGAT	1719
Db	561	GlnArgAspGlyAspAsnLysGlyAlaProLeuSerTyrThrProGluHisMetValAsn	580

1720 GCGAACTGAACGAGATGCGAGATGCGAGAGAGTGGCATGCGTGGTGGCCGCTTATCGC 1779
Db 581 AlalysLeuAsnTrpGlnIleThrGluValAlaSerTrpLeuGlyAlaArgTyrArg 600
Qy 1780 GGGAAACACACAGCTTTCACCCAGAAATATTGCTCACTGAGCGCTGTACAGAAAGAGTG 1839
Db 601 GlyJysThrProArgPheThrGlnAsnTyrSerSerLeuSerAlaValGlnLysVal 620
Qy 1840 TATGATGAGAAAGAGAAATACCTGAAAGCCCTGGAGCGGTGGATGCGAGTCTGTGCTGG 1899
Db 621 TyrAspGluLysGlyGluTyrLeuLysAlaTrpThrValValAlaGlyLeuSerTrp 640
Qy 1900 AAGATGACGATGCGCTGAGCTGCAATGCTGCGTGAATACCTGCTCAACAGAGATTAC 1959
Db 641 LysMetThrAspAlaLeuThrLeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyr 660
Qy 1960 AGTGACGTGACGCTGTACAGTCCCGTAAAGATGACGTGATGCGCGTGAATCTTCCAG 2019
Db 661 SerAspValSerLeuTyrSerAlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGln 680
Qy 2020 ACGGATCATCAACAACAGAGATGTGATACCTGAGCGAAATTAATGATGCTGCTGAAC 2079
Db 681 ThrGlySerSerThrThrGlyTyrValIleProGluArgAsnTyrTrpMetSerLeuAsn 700
Qy 2080 TATCAGTTC 2088
Db 701 TyrGlnPhe 703

RESULT 5
PCT-US95-06994-6
Sequence 6, Application PC/TUS9506994
GENERAL INFORMATION:
APPLICANT: Children's Hospital & Medical Center
APPLICANT: University of Washington
APPLICANT: Washington State University Research Foundation
APPLICANT: TARR, PHILLIP I
APPLICANT: BILGE, SIMA S
APPLICANT: BESSER, THOMAS E
APPLICANT: VARY JR, JAMES C
TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESS: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: WA 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: 07-JUN-95
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/265,714
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BRODERICK, THOMAS F
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: CHOR-18591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682 8100
TELEFAX: (206) 224 0779
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

DESCRIPTION: E. coli O157:H7 adhesin amino acid sequence,
wherein "Xaa" residues represent gaps
introduced to facilitate best alignment with
SEQ ID NO:7.
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli O157:H7
STRAIN: 86-24 NALR
PCT-US95-06994-6
Alignment Scores:
Pred. No.: 0 Length: 718
Score: 3557.00 Matches: 695
Percent Similarity: 96.80% Conservative: 0
Best Local Similarity: 96.80% Mismatches: 1
Query Match: 95.31% Indels: 22
DB: 5 Gaps: 6
US-10-625-972-4 (1-2091) x PCT-US95-06994-6 (1-718)
Qy 1 ATCGGAATAACCACTCTGGCTTCCGTAGTCATCC-----TGTCTCGGATTT 48
Db 1 MetArgIleThrThrLeuAlaSerValValIlePro*****CysLeuGlyPhe 20
Qy 49 TACCCAGCAGCATAGCTCTCTCAGAGGAT-----GTGATGATTGTC 90
Db 21 SerAlaSerSerIleAlaAlaGluAsp*****ValMetIleVal 40
Qy 91 TCGGCATCCGCTATGAGAAAGCTGACTAAACGACGCGCCAGTGTCTTCTGTGATTAC 150
Db 41 SerAlaSerGlyTyrGluLysLysLeuThrAsnAlaAlaAlaSerValSerValIleSer 60
Qy 151 CAGGAGAAATTGAGTCCAGCCAGATACACAGATGCTGGCGAGGCTCTGAGATCAGTAG 210
Db 61 GlnGluGluLeuGlnSerSerGlnTyrHisAspLeuAlaGluLeuArgSerValGlu 80
Qy 211 GGTGTGATGTTGAAAGTGTACGGGTAAACCGAGGGCTGGAATCAGCATCCGAGGA 270
Db 81 GlyValAspValGluSerGlyThrGlyLysThrGlyLysLeuGluIleSerIleA-GGly 100
Qy 271 ATCCAGCCAGTGTACAGCTGTATGATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 330
Db 101 MetProAlaSerTyrThrLeuIleLeuIleAspGlyValArgGlnGlyGlySerSerAsp 120
Qy 331 GTGACTCCCAACGGTTTTCTGCCATGAATACCGGTTTCATGCCCTCTGCGCGCCATT 390
Db 121 ValThrProAsnGlyPheSerAlaMetAsnThrGlyPheMetProLeuAlaAlaIle 140
Qy 391 GAGCGTATTGAGGTATTACAGGGGCGGATGTCCACACTGTATGCTCTGTATGCGATGGC 450
Db 141 GluArgIleGluValIleArgGlyProMetSerThrLeuTyrGlySerAlaMetGly 160
Qy 451 GGTGTGTTGAATATCATTAACCAAGAAATGACAGAAATGGCTCTCTCCGTCATGCA 510
Db 161 GlyValValAsnIleIleThrArgLysAsnAlaAspLysTrpLeuSerSerValAsnAla 180
Qy 511 GGGTGAATCTGAGGAAACCAATGGGTAAACAGCAGCCAGGTTTATTTCTGAGAC 570
Db 181 GlyLeuAsnLeuGlnLysSerAsnLysTrpGlyAsnSerSerGlnPheAsnPheTrpSer 200
Qy 571 AGTGGTCCCTTGTGGATGATTCTGTACGCTCAGGTCAGGTCAGGTCAGGTCAGGTCAG 630
Db 201 SerGlyProLeuValAspAspSerValSerLeuGlnValArgGlySerThrGlnArg 220
Qy 631 CAGGTTTCATCGGTC-----ACATCACTGAGGATACACAGCAGCGCTATTCTTAT 684
Db 221 GlnGlySerSerVal*****ThrSerLeuSerAspThrAlaGlyThrArgIleProTyr 240
Qy 685 CCCACGAGTGCACAGATTTATATCTTGTGTCACGTTTCTGCTGGAAGGCGTGGAGCAG 744
Db 241 ProThrGluSerGlnAsnTyrAsnLeuGlyAlaArgLeuAspTrpLysAlaSerGluGln 260
Qy 745 GATGTGCTCTGGTTTGTATGATGATACCCCGGAGCGGTTATGATACCGGATGGGCAA 804

Db 261 AspValLeuTrpPheAspMetAspThrThrArgGlnArgTyrAspAsnArgGlyGln 280
QY 805 CTGGGAGTCTGACGGGGGATATGACCGGACCTCGCTGATGAGCGAACAATAATTTCA 864
Db 281 LeuGlySerLeuThrGlyGlyTyrAspArgThrLeuArgTyrGluArgAsnLysIleSer 300
QY 865 GCTGGCTATGATCATATCTTCCCTTCGGAACATGGAATCTGTAACCTGGAACGAG 924
Db 301 AlaGlyTyrAspGlyThrPheThrPheGlyThrTrpLysSerTyrLeuAsnTrpAsnGlu 320
QY 925 ACAGAAATAAAGCTCGTACGCTTGACGACGAGTACTGAAGCGGCACAAATGGGGCTT 984
Db 321 ThrGluAsnLysGlyArgGluLeuValArgSerValLeuLysArgAspLysTrpGlyLeu 340
QY 985 GCCGGTCAGCCGGGAGCTTAAGGAATCGAACCTTATCTGAATTCATTCTGCTTACC 1044
Db 341 AlaGlyGlnProArgGluLeuLysGluSerAsnLeuLeuLeuAsnSerLeuLeuLeuThr 360
QY 1045 CCTCTGGGAGAACTCTCATCTGTTACGTTGGGGGGCGAGTTTCAGAGCTCGTCCATGAA 1104
Db 361 ProLeuGlyGluSerHisLeuValThrValGlyGlyGluPheGlnSerSerMetLys 380
QY 1105 GACGGAGTGTCTTCCACGACAGAGTAAACTTTCGGCAGAAAAGCTGTCGTATTT 1164
Db 381 AspGlyValValLeuAlaSerThrGlyGluThrPheArgGlnLysSerTrpSerValPhe 400
QY 1165 GCTGAGGATGAGTGGCATCTCAGGATGCACCTGCTGCTGCGGAGCGGCGCTATGAA 1224
Db 401 AlaGluAspGluTrpHisLeuThrAspAlaLeuAlaLeuThrAlaGlySerArgTyrGlu 420
QY 1225 CATCATGACAATTCGGGGACACTTCAGTCCGCTGCTATPCTGCTGGGATGTGCA 1284
Db 421 HisHisGluGlnPheGlyGlyHisPheSerProArgAlaTyrLeuValTrpAspValAla 440
QY 1285 GATGCTGAGCCTGAAAGCGGTGTGACCGGATATAGGCACCCAGATGGCGAG 1344
Db 441 AspAlaTrpHisLeuLysGlyValThrThrGlyTyrLysAlaProArgValGlyGln 460
QY 1345 CTACATAAGGATAGTGTGTGTCGGGGAGGGGAAACAAATCTACTTGGTAACCCC 1404
Db 461 LeuHisLysGlyIleSerGlyValSerGlyGlnGlyLysThrAsnLeuLeuGlyAsnPro 480
QY 1405 GACCTGAACCCGAGAGAGCTCAGTTATGAGCTGGGTGTATTAGATAACCCGCC 1464
Db 481 AspLeuLysProGluGluSerValSerTyrGluAlaGlyValTyrTyrAspAsnProAla 500
QY 1465 GGTCTGAATGCCAATGTCACAGGTTTATGACTGCTCTCCAAACAGATT-----GTC 1518
Db 501 GlyLeuAsnAlaAsnValThrGlyPheMetThrAspPheSerAsnLysIle*****Val 520
QY 1519 TCTTATTCATAAATGATACACCAATAGCTATGTAACACGCGGAAAGGCC----- 1569
Db 521 SerTyrSerIleAsnAspAsnThrAsnSerTyrValAsnSerGlyLysAla***** 540
QY 1570 CGTTTCGCGGTGGAAATTTCCCGGC-----ACATTCGCGCTGTGTCA 1614
Db 541 ArgLeuHisGlyValGluPheAlaGly*****ThrLeuProLeuTrpSer 560
QY 1615 GAGGATGTCACCTCTCATTAATTCACCTCGACCCGAGTGAACACGCTGATGTCAT 1674
Db 561 GluAspValThrLeuSerLeuAsnTyrThrTrpArgSerGluGlnArgAspGlyAsp 580
QY 1675 AACAAAGTGGCCGCTCAGTTATACCCCTCAACACATGGTGAATCGGAACCTCACTGG 1734
Db 581 AsnLysGlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrp 600
QY 1735 CAGATCACCGAAGAGTGGCATCATGCTGGGTGCCGTTATCGCGGAAACACCACTGT 1794
Db 601 GlnIleThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgGlyLysThrProArg 620
QY 1795 TTCACCCAGAAATTATCTGCTGAGCGCTGTACAGAGAAGTGTATGATGAGAAAGGA 1854

Db 621 PheThrGlnAsnTyrSerSerLeuSerAlaValGlnLysLysValTyrAspGluLysGly 640
QY 1855 GAATACCTGAAAGCTCGACGGTGGATGCGAGTCTGTCTGCGAAGATGACCGATGCC 1914
Db 641 GluTyrLeuLysAlaTrpThrValValAspAlaGlyLeuSerTrpLysMetThrAspAla 660
QY 1915 CTGACGCTGATGCTGCGGTGAATACCTGCTCAACAAGGATTACAGTACGCTGACGCTG 1974
Db 661 LeuThrLeuAsnAlaAlaValAsnLeuLeuAsnLysAspTyrSerAspValSerLeu 680
QY 1975 TACAGTCCCGGTAAAGTACGCTGTATGCGCGTGATTACTCCAGACGGGATCATCAACA 2034
Db 681 TyrSerAlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThr 700
QY 2035 ACAGGATATGATACCTGAGCGAATTAATTAATTAATTAATTAATTAATTAATTAAT 2088
Db 701 ThrGlyTyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe 718
RESULT 6
PCT-US95-06994-7
; Sequence 7, Application PC/TUS9506994
; GENERAL INFORMATION:
; APPLICANT: Children's Hospital & Medical Center
; APPLICANT: University of Washington
; APPLICANT: Washington State University Research Foundation
; APPLICANT: TARR, PHILLIP I
; APPLICANT: BILGE, SIMA S
; APPLICANT: BESSER, THOMAS E
; APPLICANT: VARY JR, JAMES C
; TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESS: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESS PLLC
; STREET: SUITE 2800, 1420 FIFTH AVENUE
; CITY: SEATTLE
; STATE: WASHINGTON
; COUNTRY: USA
; ZIP: WA 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: 07-JUN-95
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BRODERICK, THOMAS F
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: CHOR-18591
; TELEPHONE: (206) 582 8100
; TELEFAX: (206) 224 0779
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 718 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: Vibrio cholerae IgA amino acid sequence,
; DESCRIPTION: wherein xaa" residues represent gaps
; DESCRIPTION: introduced to facilitate best alignment with
; DESCRIPTION: SEQ ID NO:6.
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: VIBRIO CHOLERAE
; PCT-US95-06994-7
Alignment Scores:

Pred. No.:	1.9e-91	Length:	718
Score:	1089.50	Matches:	254
Percent Similarity:	51.53%	Conservative:	117
Best Local Similarity:	35.28%	Mismatches:	294
Query Match:	29.19%	Indels:	55
DB:	5	Gaps:	13

US-10-625-972-4 (1-2091) x PCT-US95-06994-7 (1-718)			
QY	40	CTCGGATTTTCAGCCAGCAGCATAGCTGCTGCAGAGGAT	-----GTG 81
DB	17	LeuMetPheSerAlaSer*****AlaPheAlaGlnAspAlaThrLysThrAspGluThr	36
QY	82	ATGATTTGCTCGGATCCGGCTATCAGAAAAGCTGACTAACGAGCGCCGAGTGTCT	141
DB	37	MetValValThrAlaAlaGlyTyAlaGlnValIleGlnAsnAlaPheAlaSerIleSer	56
QY	142	GTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCACCATCTGGCGAGGCTCTGAGA	201
DB	57	VallIleSerArgGluAspLeuGluSerArgTyTyArgAspValThrAspAlaLeuLys	76
QY	202	TCAGTAGAGGTGGATGTTGAAGTGGTACCGGTAACCGGAGGCTGGAAATCAGC	261
DB	77	SerValProGlyValThrVal*****ThrGlyGlyAspThrThrAspIleSer	96
QY	262	ATCCGAGGAATGCCAGCCAGTACACGCTGATCTGATGATGTTGTCGTCAGGCGGA	321
DB	97	IleArgGlyMetGlySerAsnTyrrhrLeuIleLeuValAspGlyLysArgGln--Thr	115
QY	322	AGCAGTACGCTGACTCCCAACGGTTTTTCT---GCCATGAATACCGGTTCAATCCGCT	378
DB	116	SerArgGlnThrArgProAsnSerAspGlyProGlyIleGluGlnGlyTrpLeuProPro	135
QY	379	CTGCGCCGCAATGACGTAATGAGTTATCAGGGGCGGATGTCACACTGTATGGCTCT	438
DB	136	LeuGlnAlaIleGluArgIleGluValIleArgGlyProMetSerThrLeuTyrglySer	155
QY	439	GATGGATGGCGGTGGTGGTGAATATCATATACCAAGAAGATGCAGAAATGCTCTCT	498
DB	156	AspAlaIleGlyValIleAsnIleIleThrArgLysAspGlnGlnThrTrpSerGly	175
QY	499	TCGTCATATGAGGCTGAATCTGCAGGAAGCAAAATGGGTAAACAGCAGCCAGTTT	558
DB	176	AsnValGlnLeuSerThrValValGlnGluAsnArgAlaSerGlyAspGluGlnSerAla	195
QY	559	AAATTTCTGAGCAGGCTCCCTGCTGGATGATCTGCAGCTGCAGCTACCGGTCAGC	618
DB	196	AsnPhePheValThrGlyProLeuSerAsp***AlaLeuSerLeuGlnValTyrglyGln	215
QY	619	ACACAACAGCGTCAGGGTTTCATCGGTCACTACATCACTGACGATACAGCAGGCGGTAT	678
DB	216	ThrThrGlnArgAspGluAspGluIleGluHis-----	226
QY	679	CCTTATCCACCGAGTCACAGAAATATATCTTGGTGCAGCTCTGATCGGAGCGTCG	738
DB	227	GlyTyrglyAspLysSerLeuArg---SerLeuThrSerLysLeuAsnTyrglnLeuAsn	245
QY	739	GAGCAGGATGCTCTGCTGTTTATGATGATACACCGCGGCGGTTATGATAACCGGAT	798
DB	246	ProAspHisGlnLeuGlnLeuGluAlaGlyValSerAlaGlnAspArgGluAsnVal	265
QY	799	GGGCAACTGGGAGTCAGCGGGGATATGACCGGACCTG-----	840
DB	266	GlyLysSerAlaGlnSerSerGlyCys*****ArgGlyThrCysSer	285
QY	841	-----CGCTATGAGGGAACAAATTTTCAGCTGGCTGATCATCATCTTCACC	888
DB	286	AsnThrAspAsnGlnTyrrArgArgAsnHisValAlaValSerHisGln-----	301
QY	889	TTCCGGAACATGAAA-----TCGTATCTGAATCGGACGAGACA	927
DB	302	---GlyAspTrpGlnGlyVal***GlyGlnSerAspThrThrLeuGlnTyrgluGluAsn	320

QY	928	GAATAAAGGTCGTGAGCTTGCTACGAGTGTACTGAAGCGGCACAAATGGGGCTTGC	987
DB	321	ThrAsnLysSerArgGlu*****	340
QY	988	GGTCAGCCCGGAGCTTAAGGAATCGAACCTTATTCGTAATTCATCTGCTTACCCCT	1047
DB	341	*****MetSerIleAspAsnThrValPheLysSerThrLeuValAlaPro	360
QY	1048	CTGGGAGATCTCATCTGTTACGTTGGGGGCGGAGTTTCAGAGCTCGTCCATGAAGAC	1107
DB	361	IleGlyGlu---HisMetLeuSerPheGlyValGluGlyLysHisGluSerLeuGluAsp	379
QY	1108	GGAGTTGTCTTCCAGCAGCAGGTGAACCT---TTCGGCAGAAAAGCTGGTGGTATTT	1164
DB	380	LysThrSerAsnLysIleSerSerArgThrHisIleSerAsnThrGlnTrpAlaGlyPhe	399
QY	1165	GCTGAGATGAGTGGCATCTCAGATGCATTCGCTGCCTGACTCCGGGAGCGCTATGAA	1224
DB	400	IleGluAspGluTrpAlaLeuAlaGluGlnPheArgLeuThrPheGlyGlyArgLeuAsp	419
QY	1225	CATCATGACCAATTCGGGGGACATTCAGTCCCGCGTCATATCTGCTCTGGGATGTGCA	1284
DB	420	HisAspLysAsnTyrglySerHisPheSerProArgValTyrglyValTrpAsnLeuAsp	439
QY	1285	GATGCTGACGCTGAAGCGGTGTGACACCGGATATAAGCACCAGATGGGCGAG	1344
DB	440	ProLeuTrpThrValLysGlyValSerThrGlyPheArgAlaProGlnLeuArgGlu	459
QY	1345	CTACATAAAGGATTAGTGTGTCCGGGAGGAGGAAACAAATCTACTGTTACCCCT	1404
DB	460	ValThrProAspTrpGlyGlnValSerGlyGly*****AsnIleTyrglyAsnPro	479
QY	1405	GACTGAAGCCGGAAGAGCGCTCATTTATGAGGTGGGTGTTATACGATAACCCCGCC	1464
DB	480	AspLeuGlnProGluThrSerIleAsnLysGluLeuSerLeuMetTyrrSerThrGlySer	499
QY	1465	GGTCTGAATGCCAATGTCACAGTTTATGACTGCTCTCCAAAGATTTGCTCTTAT	1524
DB	500	GlyLeuAlaAlaSerLeuThrAlaPheHisAsnAspPheLysAspLysIleThrArgVal	519
QY	1525	TCCATAAATGATAAC-----ACCAATAGCTAT	1551
DB	520	AlaCysProAlaAsnIleCysThrAlaGlyProAsnGlnTrpGlyAlaThrProThrTy	539
QY	1552	---GTAAACAGCGGAAGCCCGCTGTCAGGTGTGGAATTTCCCGCACATTCGCGCTG	1608
DB	540	ArgValAsnIleAspGluAlaGluThrTyrglyAlaAlaThrLeuSerLeuProIle	559
QY	1609	TGCTCAGAGGATGTCAGCTGTCTACCTGAATACACCTGACCCGAGTGAACACGTGAT	1668
DB	560	---ThrGluSerValGluLeuSerSerSerTyrrThrTyrrHisSerGluGlnLysSer	578
QY	1669	GGTGATAAACAAGGTGGCGCTGAGTTATACCCCTGAACACATGTTGAATGCGAACTG	1728
DB	579	GlyAsnPheAlaGlyArgProLeuLeuGlnLeuProLysHisLeuPheAsnAlaAsnLeu	598
QY	1729	AACCTGCAGATCACCGAGGTGGCATCATGCTGGTCCCGCTTATCCGCGAAACA	1788
DB	599	SerTrpGlnThrThrArgLeuAsnSerTrpAlaAsnLeuAsnTyrrArgGlyLys***	618
QY	1789	CCAGTTTCCACCCAGAAATTTCTGCTCAGTGGCGCTGTACAGAGAAGATGTATGATGAG	1848
DB	619	*****GluMetGlnProGluGlyGlyAlaSerAsnAspPheIleAlaPro*****	638
QY	1849	AAAGGAGAAATACCTGAAGCTGGACGCTGGATGCAGCTGCTGCTGCTGGAAGATGACG	1908
DB	639	*****SerTyrrThrPheIleAspThrGlyValThrValThrAlaLeuThr	658
QY	1909	GATCCCTGAGCTGGAATGCTGGGTGAATACTGCTCAACAGGATTTACAGTGCAGTG	1968
DB	659	AspThrAlaThrIleLysAlaAlaValTyrrAsnLeuPheAspGlnGlu*****	678
QY	1969	AGCCTGTACAGTCCCGGTAAAGATGACGCTGTATGCTGCTGCTGCTTCCAGACGGATCA	2028


```

Db      679 *****ValAsnTyrAlaGluTyr***** 698
QY      2029 TCAACACAGGATATGTGATACCTGAGCGAATTAATCTGATGTCGCTGACATATCATGTTTC 2088
Db      699 *****GlyTyrValGluAspGlyArgTyrTrpLeuGlyLeuAspIleAlaPhe 718

RESULT 7
US-08-765-081-6
; Sequence 6, Application US/08765081
; Patent No. 5798260
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible/Pentium
; OPERATING SYSTEM: MS-Windows 3.1
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/765,081
; FILING DATE: March 26, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997
; REFERENCE/DOCKET NUMBER: CHOR-1-10286
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: Vibrio cholerae IrgA amino acid sequence
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Vibrio Cholerae
US-08-765-081-6

Alignment Scores:
Pred. No.:      4,57e-88      Length:      652
Score:          1052.50      Matches:     258
Percent Similarity: 51.94%      Conservative: 117
Best Local Similarity: 35.72%      Mismatches: 222
Query Match:      28.20%      Indels:     125
DB:              1          Gaps:         24

US-10-625-972-4 (1-2091) x US-08-765-081-6 (1-652)
QY      40 CTCGGATTTCACCGACGACGATCT-----GCTGCAGAGGATGTGATGTT 87
Db      17 LeuMetPheSerAlaSerAlaPheAlaGlnAspAlaThrLysThrAspGluThrVal 36
QY      88 GTCCTCGGATCCGGTATGAGAAAAGCTGACTAACGCGCCGCGCAGTGTTCGTGATT 147
Db      37 ValThrAlaAlaGlyTyrAlaGlnValIleGlnAsnAlaProAlaSerIleSerValIle 56
QY      148 AGCCAGGAGGAATTCAGTCCAGCCAGTACCACGATCGCGGAGGCTCTGAGATCAGTA 207

```

```

Db      57 SerArgGluAspLeuGluSerArgTyrTyrArgAspValThrAspAlaLeuLysSerVal 76
QY      208 GAGGGTGTGATGTTGAAAGTGTGTACGGGTAAACCGGAGGGCTGGAAATCATCAGATCCGA 267
Db      77 ProGlyValThrValThrGlyGlyAspThrThr-----AspIleSerIleArg 93
QY      268 GGAATGCCAGCCAGTTACAGCTGATCTGATTGATGGTGTTCGTCTCAGGCGCGAAGCAGT 327
Db      94 GlyMetGlySerAsnTyrThrLeuIleLeuValAspGlyLysArgGln---ThrSerArg 112
QY      328 GACGTGACTCCCAACGGTTTTTCT---GCCATGAATACCGGTTTCATGCCCTCTCGGCC 384
Db      113 GlnThrArgProAsnSerAspGlyProGlyIleGluGlnGlyTrpLeuProLeuGln 132
QY      385 GCATTTCAGGGTATTGAGGTATCAGGGGCCCATGTCTCCACACTGTATGGCTGTGATGCG 444
Db      133 AlaIleGluArgIleGluValIleAe-glyProMetSerThrLeuTyrGlySerAspAla 152
QY      445 ATGGCGCGTGTGGTGAATATCATTAACAGAAAAGATGCAGACAAATGGCTCTCTCCGTC 504
Db      153 IleGlyGlyValIleAsnIleIleThrArgLysAspGlnGlnGlnIleTrpSerGlyAsnVal 172
QY      505 AATGCAGGGCTGAATCTGCAGGAAGCAACAATGGGTAAACAGCAGCCAGTTTATTTTC 564
Db      173 GlnLeuSerThrValValGlnGluAsnArgAlaSerGlyAspGluGlnSerAlaAsnPhe 192
QY      565 TGGAGCAGTGTCTCCCTTTGTGGATGATCTGTCTCAGCTGCAGGTACGCGGTACACACAA 624
Db      193 PheValThrGlyProLeu---SerAspAlaLeuSerLeuGlnValTyrGlyGlnThr 211
QY      625 CAGGTCAG-----GGTTCATCGTCTCATCATCAGTACG 657
Db      212 GlnArgAspGluAspGluIleGluHisGlyTyrGlyAspLysSerLeuArgSerLeu--- 230
QY      658 GATACAGCAGGCGCGGTATTCTTCATCCACGAGCTCACAGATTAATAATCTTGT--- 714
Db      231 -----ThrSerLysLeuAsnTyrGlnLeuAsnPro 240
QY      715 -----GCAGTCTTACTGAGGCGTGGAGCAGGATGTGTCTGGTTGATATG 765
Db      241 AspHisGlnLeuGlnLeuGluAlaGlyValSerAlaGlnAsp----- 254
QY      766 GATACCACCGCGCAGCGTTATGATAACCGGATGGGCACTGGGAGTCTGACGGGGGA 825
Db      255 -----ArgGluAsnAsnValGlyLysSerAlaGlnSerSerGlyCysArgGlyThr 271
QY      826 TATGACCGGACCC---CTGGCTATGACGGAACAAATTTTCAGCTGGCTATGATCATCT 882
Db      272 CysSerAsnThrAspAsnGlnTyrArgArgAsnHisValAlaValSerHisGln----- 289
QY      883 TTCACCTTCGGAACATGGAAA-----TCGTATCTGAACCTGGACACGAG 924
Db      290 -----GlyAspTrpGlnGlyValGlyGlnSerAspThrTyrLeuGlnTyrGluGlu 306
QY      925 ACAGAAAATAAAGTCTCGTCTGAGCTTTGACGAGTGTACTGAAGCGCGACAAATGGGGCTT 984
Db      307 AsnThrAsnLysSer----- 311
QY      985 GCGGTCACCGCGGAGGAGCTTAAGGAATCGAACCTTATCTGAAATTCATTCTGCTTACC 1044
Db      312 -----ArgGluMetSerIleAspAsnThrValPheLysSerThrLeuValAla 327
QY      1045 CCTCTCGGAGAAATCTCATCTGTTACGGTGGGGCGGAGTTTCAGAGCTCGTCCATGAAA 1104
Db      328 ProIleGlyGlu---HisMetLeuSerPheGlyValGluGlyLysHisGluSerLeuGlu 346
QY      1105 GACGGAGTTGTCTTCCAGCAGCAGTGAACCT---TTCGGCGCAAAAGCTGTGCGTA 1161
Db      347 AspLysThrSerAsnLysIleSerSerArgThrHisIleSerAsnThrGlnTrpAlaGly 366
QY      1162 TTTCGTGAGGATGAGTGGCATCTCAGGATGCACCTTGGCTGACCTGCGGCGACGCGCTAT 1221

```


Alignment Scores:		6,41e-84	Length:	677
Pred. No.:	Score:	1007.50	Matches:	252
Percent Similarity:		50.74%	Conservative:	123
Best Local Similarity:		34.10%	Mismatches:	237
Query Match:		27.00%	Indels:	127
DB:		4	Gaps:	23
US-10-625-972-4 (1-2091) x US-09-543-681A-7936 (1-677)				
QY	16	CTGGCTCCGCTAGTCACTCTCTCGGATTTTCAGCCAGCAGCATAGCTGCTGCAGAG	75	
DB	18	ILeAlaLeuGlyValValAlaAlaLeuAlaSerSerPheValMetThrAlaSerAlaGlu	37	
QY	76	GAT-----GTGATGATTGCTCGGCATCGGCTATGAGAAAAGCTGACTACGCA	126	
DB	38	AspProGluArgLeuValThrThrAlaSerGlyPheLysGlnThrValGluAspAla	57	
QY	127	GCAGCAGTCTTCTGATTAGCAGGAGCAATTGCAGTCCAGCCAGTACACGATCTG	186	
DB	58	ProLysSerValSerValThrArgGluGlnLeuGluThrLysSerTyrArgAspVal	77	
QY	187	GCAGAGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGGTACGGTAAACCGGA	246	
DB	78	ThrAspAlaLeuLysAspValProGlyValLeuVal-----ThrGlyGlySer	94	
QY	247	GGGTGGAATCAGATCCGAGGAATCCAGCCAGTTCACAGCTGATACCTGATGATGGT	306	
DB	95	SerSerAspLysSerIleArgGlyMetAspProLysTyrThrMetIleLeuValAspGly	114	
QY	307	GTTCGTGAGGCGGAAGCAGTGCAGTGCCTCCCAACCGT---TTTCTGCCATGAATACC	363	
DB	115	LysArgVal---AlaSerArgGluThrArgProAsnSerAspAsnSerGlyIleGluGln	133	
QY	364	GGGTTCATGCCCCCTCTGGCCGATTCAGGGTATTCAGCTATCAGGGGCGGATCTCC	423	
DB	134	GlyTrpLeuProLeuProAlaIleGluArgIleGluValArgGlyProMetSer	153	
QY	424	ACACTGTATGCTCTGATCGCATGGCGGTGTGGATATCATTCACAGAAAAGATGCA	483	
DB	154	SerLeuTyrGlySerAspAlaMetGlyGlyValIleAsnIleThrArgLysAlaGln	173	
QY	484	GACAAATGGCTCTTCGCTCAATGCAGGGCTGAATCTGCAGGAAACCAATGGGGT	543	
DB	174	LysGluTrpAsnPheSerLeuArgGlyAspThrThrLeuThrGluArgLysAsnGluGly	193	
QY	544	AACAGCAGCAGTTAATTTCTGGAGCAGTGCCTCTGTGGATGATTCGTGAGCTG	603	
DB	194	AsnThrGlyGlnGlySerPheTyrAlaAlaGlyProLeuIleAspAsnValLeuGlyLeu	213	
QY	604	CAGGTACCGGTAGCACACACAGCGTCAGGGTTCATCGGTTCATCAGTACGAGTACA	663	
DB	214	LysLeuGlnGlyGlnTyrSerHisArg-----Gly	223	
QY	664	CGAGCAGCGGTATCTCTATCCACGAGTCACAGAAATTAATCTTGGTGCAGCTT	723	
DB	224	GluAspArgArgValAspGlyTyrAsnArgGlnIleThrAlaSerGlyGlyLeu	243	
QY	724	GACTGGAAGCGTCCGACAGAGTGTCTGTGGTTTGATATGATACACCCCGCAGCGT	783	
DB	244	SerTrpThrProAspGlyLysAsnThrValGluPheGluPheLysAspAsnGlnHis	263	
QY	784	TATGATAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGACCTCGGC	843	
DB	264	ArgAspSerArg-----ValGlyTyrSerLysSerAlaGlu	275	
QY	844	-----TATGCGGAAACCAAAATTTCA	864	
DB	276	LeuValLysGlyLysAlaProGlnSerSerPheThrAspTyrGluLeuThrHisTyrAla	295	
QY	865	GCTGCTATGATCATCTTACCTTCGGAACATCGAATCGATCTGAATCGAACGAG	924	

DB	296	LeuThrHisAspGlyIleTyrAspPheGlyThrMetAsnThrTyrValGlnArgAspGlu	315	
QY	925	ACAGAAATTAAGCTCGTGCAGCTTGTAGCGCAGTCTACTGAAGCGCGCAAAATGGGGCTT	984	
DB	316	SerArgAsnProSerArg-----LysMetAsp-----	324	
QY	985	CGCGTTCAGCCGCGGAGCTTAAAGAAATCGAACTTATCTGGAATTTCTACTGCTTACC	1044	
DB	325	-----TyrAspAspThrThrValArgAsnGlnThrValPhe---	336	
QY	1045	CCTCTGGGAGAAATCTCATCTGTTACGGTGGGCGAGTTCACAGCTCTCTCATGAAA	1104	
DB	337	---MetPheGlyAspHisMetLeuSerValGlyGlyGlnTyrArgTyrGluGluLeuLys	355	
QY	1105	GAC---GGAGTGTCTCTCCAGCAGCAGGTGAACTTTCGCGCAGAAAAGCTGTGCGTA	1161	
DB	356	AspGluGlyAsnLysLeuLysGlyAsnLysLeuAspArgTyr---SerTrpAlaLeu	374	
QY	1162	TTTGTGAGGATGAGTGCATCTCAGCATGCATCTCGCTGCGTGCATATCTGCTCGGATGTG	1221	
DB	375	PheAlaGluAspGluTrpThrLeuThrAsnAspPheAlaLeuThrAlaGlyLeuArgMet	394	
QY	1222	GAACATCATGACCAATTCGGGGGACACTTCAGTCCGCTGCATATCTGCTCGGATGTG	1281	
DB	395	AspLysAspGluAsnPheGlyThrHisTrpThrProArgValTyrGlyValTrpHisLeu	414	
QY	1282	GCAGATGCTGCACCTGAAAGCGGTGTGACCGGATATAGCCAGCCAGATGGG	1341	
DB	415	AlaAspGluTrpThrLeuLysGlyValSerThrGlyTyrArgSerProAspLeuArg	434	
QY	1342	CAGCTACATAAGGATAGTGTGTGTCGGGGCAGGAGGAAAACAAAT-----CTACTT	1395	
DB	435	GlnAlaThrAlaThrTrpGlyGlnGlyThrGlyGlyArgTyrAspAlaValIleTyr	454	
QY	1396	GGTAACCCCGCCTCAAGCCGGAAGAGCGCTCAGTTATGAGGCTGGGTGATTCAGT	1455	
DB	455	GlyAsnProAspLeuLysProGluLysSerValThrGluGluIleGlyIleTrpAsp	474	
QY	1456	AACCCCGCGCTGATGCCAATCTCAGGTTTATGACTGCTTCCCAACAGATT	1515	
DB	475	AsnArgAspAsnLeuThrAlaSerLeuThrIleTyrAsnThrAspPheLysAspLysIle	494	
QY	1516	GTCTCTTAT-----TCCATAAATGATAACACCAATAGCTATGTA-----	1554	
DB	495	ThrGluValArgArgCysAspSerArgSerAspLysAlaProSerMetLeuAspCysThr	514	
QY	1555	-----AACAGCGGA-----	1563	
DB	515	LeuAlaAspGlyValGlyAsnSerGlyAspProTyrAspPheValSerAspArgThrAsn	534	
QY	1564	-----AAGCCCGGTTCACCGTGTGGAATTTGCGGCACATTCGCGCTGTGG-----	1611	
DB	535	ValAspLysAlaAsnMetArgGlyValGluValThrAlaAsn-----TrpIleIle	551	
QY	1612	TCAGAGATGTACGCTGTCATCTGAATTACCTGGACCCGAGAGTGAACACGTGTGCT	1671	
DB	552	SerProGluTrpAsnLeuAlaAsnTyrThrPheThrAspThrGluGlnLysSerGly	571	
QY	1672	GATAACAAAGGTCCGCGCTGATATACCTCGAACACATGCTGAATCGAAACTGAAC	1731	
DB	572	AspPheLysGlyLysProLeuAsnLysGlnProArgHisMetAlaAsnAlaThrLeuAsn	591	
QY	1732	TGGCAGATCACCGAAGGTGGCATCATGCTCGCTGGTCCCGTTCATCGCGGAAAAACCA	1791	
DB	592	TrpGluThrThrProGluMetGluThrTrpAlaArgIleAsnPheArgGlyLys-----	609	
QY	1792	CGTTTCCACCAAGATTTCTGCTACGCGCTGTACAGAGAAAGTGTATGATGAGAAA	1851	
DB	610	-----ThrSerAspTyrLeuSerArgThrSerMetSerLysSerThr-----	623	
QY	1852	GGAGANTACCTGAAAGCTGCAGCGTGTGGATGACAGTCTGCTGCGGAGATGACCGAT	1911	
DB	624	-----ProSerTyrAlaPheValAspValGlyThrSerTyrSerLeuThrLys	639	

1528 ATAAATGAT-----AACACCAATAGCTATGTAACACGCGAAAG 1566
QY
Db 511 ThrGlycylLeuAspProIleThrGlyLeuLysLeuTyrGlnTyrAspAsnValGlyLys 530
QY 1567 GCCGGTGGCAGGTGGGAATTCGCCGGCACATTCGCCGTGGTCAGAGGATGTCACG 1626
Db 531 AlaAsnIleLysGlyIleGluThrAlaValAlaPheProVal---AlaAspAsnIlePheArg 549
QY 1627 CTGTCACCTGAATACACCTGGACCCGAAAGTGAACACGCTGATGCTGATAC----- 1677
Db 550 ValSerAlaAsnTyrThrTyrIleAsnSerLysArgLysSerAspAspGluLysLeuGly 569
QY 1678 -----AAAGTGGCCCGCTGAGTTATACCCCTGAACACATGGTGAATGG 1722
Db 570 SerGlyGluSerLeuLysGlyTyrProLeuAspMetThrProLysHisSerAlaAsnAla 589
QY 1723 AAACGTGAACCTGGCAGATCAGCAGAGGTGGCATGCTGGTGGCCCGTTCGCGGG 1782
Db 590 ArgValAspIleThrGlyTyrAspGluAlaThrSerPheTyrAlaAsnThrAlaTyrThrGly 609
QY 1783 AAAACACACACCGTTTCAACCCAGAAATTTATCGTCACTAGCGCTGTACAGAAAGAGTGTAT 1842
Db 610 Lys -----GlnIleThrAlaAlaGlnArgAsnGlyTyr 620
QY 1843 GATCAGAAAGAGAGATACCTGAACCTGGACGCTGGTGGTGCAGCTGTGCTGGAAG 1902
Db 621 ---ThrGlyAlaArgTyrArgSerGlyTyrThrThrPheAspLeuGlyMetThrTyrAsn 639
QY 1903 ATGACCGATGCCCTGACCGCTGAATGCTGGGTGAATTAACCTGTCAACAGAGATACAGT 1962
Db 640 PheAsnLysAsnThrMetLeuAsnLeuAlaValLeuAsnIleThrAsp----- 655
QY 1963 GACGTGACCTGTACAGTGGCGGTGAAGTACGCTGTATGCCGCTGATCTTCCAGACG 2022
Db 656 -----GluThr 657
QY 2023 GGATCATCAACAACA-----GGATATGCTGATACCTGAG---CGAAATTAATCTGG 2067
Db 658 GlyProAlaValAsnAspLysGlyValAsnTrpValValAspGluGlyArgArgTyrTrp 677
QY 2068 ATGTCGTGAACCTACAGTTC 2088
Db 678 AlaAsnIleLysTyrSerPhe 684

RESULT 11
US-09-489-039A-8133
; Sequence 8133, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8133
; LENGTH: 699
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8133

Alignment Scores:
Pred. No.: 3,138-72 Length: 699
Score: 880.50 Matches: 236
Percent Similarity: 49.73% Conservative: 123
Best Local Similarity: 32.73% Mismatches: 249
Query Match: 23.53% Indels: 113
DB: 4 Gaps: 21

US-10-625-972-4 (1-2091) x US-09-489-039A-8133 (1-699)
QY 43 GGATTTTCAGCAGCAGCATAGCTGCTGCA-----GAGGAT 78
Db 53 GlyLeuSerAlaSerValValSerLeuAlaPheProAlaLeuAlaAspValAsnGluGlu 72
QY 79 GTGATGATTGCTCGGCTCCGGCTATGAGAAAAGCTGACTAAACGACGAGCCAGCTGTT 138
Db 73 ThrLeuValValThrAlaSerAlaThrGluGlnAsnValLysAspAlaProAlaSerIle 92
QY 139 TCTGTGATTAGCAGGAGAAATTCAGTCCAGCCAGTACCAGTCTCGGCGAGCTCTG 198
Db 93 SerValIleThrGlnAspLeuGlnArgLysProValGlnAsnLeuLysAspValLeu 112
QY 199 AGATCAGTAGAGGTGCTGATGTT---GAAAGTGTACGGGTAAACCGGAGGGCTGAA 255
Db 113 ArgAspValProGlyValGlnLeuThrAsnGluGlyAspAsnArgLysGly----- 129
QY 256 ATCAGCATCCGAGGAAATCCAGCCAGTTCACCGCTGATACGATTGATGGTGTTCGTGAG 315
Db 130 ValSerIleArgGlyLeuSerSerTyrThrLeuIleLeuValAspGlyLysArgVal 149
QY 316 GCGGAGACGATGACGTGATCCCAACGGTTTTCTGCCATGAATACCGGGTTCATGCC 375
Db 150 AsnSerArgAsnAlaValPheArgHisAsnAspPheAspLeuAsn-----TrpIle 166
QY 376 CCTCTGGCCCGCATTGAGCGTATTGAGGTATTACAGGGGCCGATGCTCCACACTGTATGCC 435
Db 167 ProValAspAlaIleGluArgIleGluValValArgGlyProMetSerSerLeuTyrGly 186
QY 436 TCTGATGCGATCGGCGGTGTGTGATATCATTCACAGAAAGATGCAGACAAATGGCTC 495
Db 187 SerAspAlaLeuGlyValValAsnIleIleThrLysLysIleGlyGlnLysTrpThr 206
QY 496 TCTTCGCTCAATGCGAGGCTGATCTGCAGGAGAAACCAACAAATGGGTAAACAGCAGCCAG 555
Db 207 GlyThrLeuSerAlaAspThrThrIleGlnGluHisArgAspArgGlyAspThrTyrAsn 226
QY 556 TTATATTTCTGGAGCAGTGTCCCTTGTGGATGATTCTGTGAGCTGACAGTGCAGGTACGGGT 615
Db 227 GlyGlnPhePheThrSerGlyProLeuIleAspGlyValLeuGlyMetLysAlaTyrGly 246
QY 616 AGCACACACAGGTCAGGTTTCATCGTCACTCATCATCAGCGAGATACACAGC----- 669
Db 247 SerLeuAlaLysArgAlaLysAspAspProGlnSerSerSerAsnAlaThrGlyGluThr 266
QY 670 ACGGCTATTCTTATCCACGAGTACACAGAAATTAATCTTGGTGCACGCTTGTGACTGG 729
Db 267 ProArgIleGluGlyPheThr---SerArgAspGlyAsnVal-----GluPheAlaTrp 283
QY 730 AAGGCTCGAGCAGGATGCTGCTGTTGATGATATGATACACCGCGGCGGCTATGAT 789
Db 284 ThrProAsnGluAsnHisAspPhe-----ThrAlaGlyTyrGlyPheAsp 298
QY 790 AACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCTCGGCTATGAG 849
Db 299 ArgGlnAspArgAspSerAspSerLeu-----AspArgAsn---ArgLeuGlu 313
QY 850 CGAAACAAATTTTCAGCTGCTATGATCATATCTTCCATTCGGAACATGGAATCGTAT 909
Db 314 ArgGluAsnTyrSerLeuSerHisAsnGlyArgTrpAspIleGlyAsnSerGluLeuLys 333
QY 910 CTGAACCTGACAGCAGACAGAAATTAAGTTCGTGAGCTTGTGACGCTGCTACTGAAGCGC 969
Db 334 PheTyrGlyGluLysValAspAsnLysAsn----- 343
QY 970 GACAAATGGGGCTTGGCGGTTCAGCGCGGAGCTT---AAGGAATCGAACCTTATCTCTG 1026
Db 344 -----ProGlyGlnSerGlyThrIleThrSerGluSerAsnAlaIleAsp 358
QY 1027 AATTCATTACTGCTTACCTCCCTCTGGAGAA---TCTCATCTGTTACGGTGGGGCGAG 1083
Db 359 GlyLysTyrValLeu---ProLeuGlyMetIleAsnGlnLeuValThrPheGlyGlyGlu 377

RESULT 13
 US-08-765-081-7
 ; Sequence 7, Application US/08765081
 ; Patent No. 5798260
 ;
 ; GENERAL INFORMATION:
 ; APPLICANT: Tarr, P.I., Balge, S.S., Besser, T.E., Vary Jr., J.C.
 ; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
 ; NUMBER OF SEQUENCES: 12
 ;
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSES: Christensen, O'Connor, Johnson and Kindness
 ; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
 ; CITY: Seattle

Alignment Scores:	6.93e-65	Length:	663
Pred. No.:	800.50	Matches:	231
Score:	50	Mismatches:	120
Percent Similarity:	41.1%	Conservative:	242
Best Local Similarity:	31.0%	Mismatches:	151
Query Match:	21.4%	Indels:	151
DB:	1	Gaps:	25

US-10-625-972-4 (1-2091) x US-08-765-081-7 (1-663)

```

40 CTCGATTTTCAGCCAGCAGCATAGCTGCTGCA----- 72
   ::::: |||::: |||
10 ValGlyLeuCysLeuSerAlaIleSerCysAlaTrpProValLeuAlaValAspAsp 29
   ::::: |||::: |||
73 GAGGATGTGATGTTGTCGCGATCCGGCTATGAGAAAAGCTGACTAACGACGCGGC 132
   ::::: |||::: |||
30 GlyGluThrMetValValThrAlaSerSerValGluGlnAsnLeuLysAspAlaPro 49
   ::::: |||::: |||
133 AGTGTTTCTGTGATTAGCCAGCAGCAATTGCAGTCCAGCCAGTACCACCATCTGGCGGAG 192
   ::::: |||::: |||
50 SerIleSerValIleThrGlnGluAspLeuGlnArgLysProValGlnAsnLeuLysAsp 59
   ::::: |||::: |||
193 GCTCTGAGATCAGTAGAGGGTGTGGATGTT---GAAAGTGGTACGGGTAAACCGGAGG 249
   ::::: |||::: |||
70 ValLeuLysGluValProGlyValGlnLeuThrAsnGluGlyAspAsnArgLysGly--- 88
   ::::: |||::: |||
```

250	CTGGAATCAGCATCCGAGGAATCCAGCCAGTACACCGCTGATCTGATTTGATGGTTT	309
Qy		
89	-----ValSerIleArgGlyLeuAspSerSerTyThrLeuIleuValAspGlyLys	106
Db		
310	CGTCAGGGCGGAACGATGACGTGCCCAACGGTTTTTCTGCCATGAATACCGGGTTC	369
Qy		
107	ArgValAsnSerArgAsnAlaValPheArgHisAsnAspPheAspLeuAsn-----	123
Db		
370	ATGCCCCCTCTGCCGCCATTAGCGCTATTGAGGTTATCAGGGGGCGGATGTCACACATG	429
Qy		
124	TrpIleProValAspSerIleGluArgIleGluValValArgGlyPrometSerSerLeu	143
Db		
430	TATGGCTCTGATCGGATGGCGGTGGTGGAAATCATTTACCAAGAAAGATCGAGACAA	489
Qy		
144	TyrGlySerAspAlaLeuGlyGlyValValAsnIleIleThrIlysIleGlyGlnLys	163
Db		
490	TGCGTCTCTCCGTCAATCGACGGCTGAATCTCGAGGAAGCAACAATGGGTAACAGC	549
Qy		
164	TrpSerGlyThrValThrValAspThrThrIleGlnGluHisArgAspArgGlyAspThr	183
Db		
550	AGCAGCTTAAATTTTCGAGCAGTGTCCCTTGTGGATGATCTCTCAGCCTCGAGGTA	609
Qy		
184	TyrAsnGlyGlnPhePheThrSerGlyProLeuIleAspGlyValLeuGlyMetLysAla	203
Db		
610	CGCGGTAGCACACAACAGCT-----CAGGTTCACTCGGTCACTCACTCAGTCG	654
Qy		
204	TyrGlySerLeuAlaIysArgGluLysAspAspProGlnAsnSerThrThrThrAspThr	223
Db		
655	AGCGATACAGCAGCGCGGTATTCTCTTATCCCAACGAGCTCACAGAATTATACTTGGT	714
Qy		
224	GlyGluThr-----	226
Db		
715	GCACGCTTCGACTGGGAAGCGTCGGAGCAG-----CATGTGCTCTGGTTTGATATG	765
Qy		
227	ProArgIleGluGlyPheSerSerArgAspGlyAsnValGluPheAlaTrp-----	243
Db		
766	GATPACCACCGGAGCGGTATTGATAACCGGGATGGGCAACTGGGGAGTCTGACGGGGGA	825
Qy		
244	--ThrProAsnGlnAsnHisAspPheThrAlaGlyTyr-----Gly	256
Db		
826	TATGACGG-----ACCTGCGCTATGAGCGAACAACA	858
Qy		
257	PheAspArgGlnAspArgAspSerAspSerLeuAspLysAsnArgLeuGluArgGlnAsn	276
Db		
859	ATTTCAGCTGGCTATGATCATCTTTCACTTCGGAACATCGTAATCTGAACTGG	918
Qy		
277	TyrSerValSerHisAsnGlyArgTAspTyrGlyThrSerGluLeuLysTyrTyrGly	296
Db		
919	AACGACACAGAAATAAAGGTCGTGAGCTGTACGCGATGCTACTGAAGCGGACAAATGG	978
Qy		
297	GluLysValGluAsnLysAsn-----	303
Db		
979	GGGCTTTCGGCTCAGCCGGCGGAGCTT--AAGGAATCGAACCTTATCTCGAATTCATTA	1035
Qy		
304	-----ProGlyAsnSerSerProIleThrSerGluSerAsnThrValAspGlyLysTyr	321
Db		
1036	CTGCTTACCCCTCTGGGAAATCTCATCTGGTTACGTCGGGGCGGAGCTTCAGAGCTCG	1095
Qy		
322	ThrLeuProLeuThrAlaIleAsnGlnPheLeuThrValGlyGlyGluThrPargHisAsp	341
Db		
1096	TCCATGAAGACGGAGTTCTCTGCCAGCACAGGTGAAACTTCCGGCAGAAAAAGC---	1152
Qy		
342	LysLeuSerAspAlaValAsnLeu-----ThrGlyGlyThrSerSerLysThrSerAla	359
Db		
1153	-----TGTCGGTATTCTCAGGATGAGTGCATCTCACGGATCACTTGGCTGACT	1206
Qy		
360	SerGlnTyrAlaLeuPheValGluAspGluThrPargIlePheGluProLeuAlaLeuThr	379
Db		
1207	GGGGCAGCGCTATGAACATCATGACGAATTCGGGGGACATTCAGTCGGGTGCATAT	1266
Qy		
380	ThrGlyValArgMetAspAspHisGluThrTyrGlyGluHisTrpSerProArgAlaTyr	399
Db		

QY 1267 CTGGCTGGGATGCGGACGCTGACGCTGAAGCGGCTGACACGGGATTAAG 1326
Db 400 LeuValTyrAsnAlaThrAspThrValThrValIysGlyGlyTrpAlaThrAlaPheLys 419
QY 1327 GCACCCAGATGGGCGACCTACATAAAGGATTAAGTGTGTCGCGGCGAGGAAACAA 1386
Db 420 AlaProSerLeuLeuGlnLeuSerProAspTrpThrSerAsnSerCysArgGlyAlaCys 439
QY 1387 AATCTACTGTGTAACCCGACCTGACCGGAGAGAGCGTCAGTTATGAGCGTGGGTG 1446
Db 440 LysIleValGlySerProAspLeuLeuProGluThrSerGluSerTrpGluLeuGlyLeu 459
QY 1447 TATTACGATAACCCG-----GCCGGTCTGAATGCCAATGTCACAGGTTTATG 1494
Db 460 TyrTyrMetGlyGluGluGlyTrpLeuGluGlyValGluSerSerValThrValPheArg 479
QY 1495 ACTGACTTCTCCACAGAGATTGCTCTTATTCATAAATGATACACCAAT----- 1545
Db 480 AsnAspValLysAspArgile---SerIleSerArgThrSerAspValAsnAlaPro 498
QY 1545 ----- 1545
Db 499 GlyTyrGlnAsnPheValGlyPheGluThrGlyAlaAsnGlyArgGlyProValPhe 518
QY 1546 AGCTATGTAAACAGCGGAAGCGCGGTGACGCTGTGGAATTTGCCGCGACATTCCCG 1605
Db 519 SerTyrTyrAsnValAsnLysAlaArgAsnGlnGlyValGluThrGluLeuLysIlePro 538
QY 1606 CTGTGCTCAGAGGATGTCACCTGCTCACTGTAATACACCTGACCGGACCGAAGTGAACAGT 1665
Db 539 ---PheAsnAspGluTrpLysLeuSerIleAsnTyrThrTyr-----AsnAspGlyArg 555
QY 1666 GAT-----GGTGATAACAAAGTGGCGCGTGTAGTTATACCCCTGAACACATGGTG 1716
Db 556 AspValSerAsnGlyGluAsnLys-----ProLeuSerAspLeuProPheHisLeuAla 573
QY 1717 AATGCGAACTGAATGCGCAGATCACCAGAGGTGGCATCATGGCTGGGTGGCGCGTTAT 1776
Db 574 -----LeuGluAspTrpSerPheTyrValSerGlyHisTyr 585
QY 1777 CGCGGAAACACACCGGTTTCAACCCAGATTTATCGTCACGTGAGCGGTGACAGAA 1836
Db 586 ThrGly-----GlnLysArg 590
QY 1837 GTGTATGATGAGAAGAGAGATACCTGAAACCTGGACGGTGGTGGATGACAGGTCTGTCG 1896
Db 591 AlaAspSerAlaThrAlaLysThrProGlyGlyTyrThrIleTrpAsnThrGlyAlaAla 610
QY 1897 TGGAGATGACGGATGCCCTGACGCTGAATGCTGGGTGAATACCTGCTCAACAGGAT 1956
Db 611 TrpGlnValThrLysAspValLysLeuArgAlaGlyValLeuAsnLeuGlyAspLysThr 630
QY 1957 TACAGTACGTGAGCGCTGTACAGTCCCGGTAAAGATACGCTGTATGCGGTGATTACTTC 2016
Db 631 AlaAsnGlyThrLeuAspTrp-----LysProAspLeuSerArgAspTyr--- 646
QY 2017 CAGCGGATCATCAACACAGGATATGTGATACCTGAGCGGAATTAAGTATGATGCTGCTG 2076
Db 647 -----SerTyrAsnGluAspGlyArgArgTyrPheMetAlaVal 659
QY 2077 AACTATCAGTTC 2088
Db 660 AspTyrArgPhe 663

RESULT 14

US-09-098-082-7
; Sequence 7, Application US/09098082
; Patent No. 6040421
; GENERAL INFORMATION:
; APPLICANT: Barr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
STREET: 2800 Pacific First Center, 1420 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
COMPUTER: IBM PC compatible/Pentium II
OPERATING SYSTEM: MS-Windows 95
SOFTWARE: Word for Windows-6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098,082
FILING DATE: June 24, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/765,081
FILING DATE: March 26, 1997
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995
APPLICATION NUMBER: US 08/265,714
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: CHOR-1-12402
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 663 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: E. coli CirA protein amino acid sequence
HYPOTHETICAL: NO
ORGANISM: Escherichia Coli
US-09-098-082-7

Alignment Scores:

Pred. No.: 6 93e-65 Length: 663
Score: 800.50 Matches: 231
Percent Similarity: 47.18% Conservative: 120
Best Local Similarity: 31.05% Mismatches: 242
Query Match: 21.45% Indels: 151
DB: 3 Gaps: 25

US-10-625-972-4 (1-2091) x US-09-098-082-7 (1-663)

QY 40 CTCGGATTTTCAGCCAGCAGCATAGCTGTGCA----- 72
Db 10 ValGlyLeuCysLeuSerAlaIleSerCysAlaTrpProValLeuAlaValAspAsp 29
QY 73 GAGATGTGTATGATTCCTCGGCATCCGGCTATGAGAAAAGCTGACTAACGCGCGCC 132
Db 30 GlyClnThrMetValValThrAlaSerSerValGlnAsnLeuLysAspAlaProAla 49
QY 133 AGTGTTCCTGTGATTACCGAGGAAATTCAGTCCAGCCAGTACCACGATCTGCGCGAG 192
Db 50 SerIleSerValIleThrGlnGluAspLeuGlnArgLysProValGlnAsnLeuLysAsp 69
QY 193 GCTCTGAGATCAGTACAGGCTGTGGATGT---GAAAGTGGTACGGTAAACCGGAGG 249
Db 70 ValLeuLysGluValProGlyValGlnLeuThrAsnGluGlyAspAsnArgLysGly--- 88
QY 250 CTGGAATCAGCATCCGAGGAATGCCAGCCAGTACACCTGATCTGATGATGTT 309
Db 89 -----ValSerIleArgGlyLeuAspSerTyrThrLeuIleLeuValAspGlyLys 106
QY 310 CGTCAGGCGGAGAGCAGTACGTCATCCCAACGGTTTTCTGCCATGATACCGGTTTC 369
Db 107 ArgValAsnSerArgAsnAlaValPheArgHisAsnAspPheAspLeuAsn----- 123

QY 370 ATGCCCTCTGCGCCGCTTATGAGCGTATAGGTTATACAGGGGCGGATGTCACACTG 429
 Db 124 TptIleProValAspSerIleGluArgIleGluValValArgGlyProMetSerSerLeu 143
 QY 430 TATGGCTCTGATCGATGGCGGTGGTGAATATCATTTACAGAAAGAAATGCAGACAAA 489
 Db 144 TyrGlySerAspAlaLeuGlyGlyValValAsnIleIleThrLysLysIleGlyGlnLys 163
 QY 490 TGCTCTCTCCCTCAATCAGCGCTGATCTGCAGGAAAGCAACAATGGGGTAACAGC 549
 Db 164 TptSerGlyThrValThrValAspThrThrIleGlnGluHisArgAspArgGlyAspThr 183
 QY 550 AGCCAGTTAAATTTGAGAGCAGTGGTCCCTGTGGATGATCTGTGAGCTGCAGGTA 609
 Db 184 TyrAsnGlyGlnPhePheThrSerGlyProIleIleAspGlyValLeuGlyMetLysala 203
 QY 610 CGCGGTAGCACACACAGCGT-----CAGGGTTTCATCGGTGCATCACTG 654
 Db 204 TyrGlySerLeuAlaLysArgGluLysAspAspProGlnAsnSerThrThrAspThr 223
 QY 655 AGGATACAGCAGGACCGGTATCTCTTATCCAGGAGTGCAGAAATTAATCTTGGT 714
 Db 224 GlyGluThr----- 226
 QY 715 GCACGTCTTGAAGCGGCGTGGAGCAG-----GATGTGCTCTGGTTTGATATG 765
 Db 227 ProArgIleGluGlyPheSerSerArgAspGlyAsnValGluPheAlaTrp----- 243
 QY 766 GATACACCGCGCAGCGTTATGATACCGGGATGGCAACTGGGGAGTCTGACGGGGGA 825
 Db 244 ---ThrProAsnGlnAsnHisAspPheThrAlaGlyTy-----Gly 256
 QY 826 TATGACCGG-----ACCGTCGCTATGACGGAACAAA 858
 Db 257 PheAspArgGlnAspArgAspSerLeuAspLysAsnArgLeuGluArgGlnAsn 276
 QY 859 ATTTCAGCTGGCTATGATCATCTTTACCTTCGGAACATCGAATCTGTAACCTGG 918
 Db 277 TyrSerValSerHisAsnGlyArgTrpAspTyrGlyThrSerGluLeuLysTy-TyrGly 296
 QY 919 AACGACACAGAAATAAGGTCGTGAGCTGTACCGAGTACTAAGCGCGCACAAATGG 978
 Db 297 GluLysValGluAsnLysAsn----- 303
 QY 979 GGGCTTGGCGGTGAGCGCGGAGCTT---AAGGAATCGAATCTATCTGTAATTCATTA 1035
 Db 304 -----ProGlyAsnSerSerProIleThrSerGluSerAsnThrValAspGlyLysTy 321
 QY 1036 CTGCTTACCCCTCTGGGAGAACTCATCTGTTAGCTGGGGGGGAGTTTCAGAGCTCG 1095
 Db 322 ThrLeuProLeuThrAlaIleAsnGlnPheLeuThrValGlyGlyGluTrpArgHisAsp 341
 QY 1096 TCATGAAAGACGAGGTCTCTCTGCCACACAGGTGAACCTTCCGGCAGAAAGC--- 1152
 Db 342 LysLeuSerAspAlaValAsnLeu-----ThrGlyGlyThrSerSerLysThrSerAla 359
 QY 1153 -----TGTCGGTATTTCTGAGATGAGTGGCATCTCAGGATGCACTTGGCTGACT 1206
 Db 360 SerGlnTyThrAlaLeuPheValGluAspGluTrpArgIlePheGluProLeuAlaLeuThr 379
 QY 1207 GCGGCGAGCGGTATCAATCATGCAATTCGCGGGGACACTTCAGTCCGCGTCATAT 1266
 Db 380 ThrGlyValArgMetAspAspHisGluThrTyGlyGluHisTrpSerProArgAlaTy 399
 QY 1267 CTGGCTGGGATGTGGCAGATCCCTGGAGCTGAAAGCGGTGTGACCAACCGGATATAG 1326
 Db 400 LeuValTyAsnAlaThrAspThrValThrValLysGlyGlyTrpAlaThrAlaPheLys 419
 QY 1327 GCACCCAGATGGGCGAGTACATAAAGGATTAAGTGGTGTCTCCGGCGCAGGAAACAA 1386
 Db 420 AlaProSerLeuGlnLeuSerProAspTrpThrSerAsnSerCysArgGlyAlaCys 439

QY 1387 AATCTACTTGTTAAACCCCGACCTGAGCGGAGAGAGCGTTCAGTTATCAGGCTGGGTG 1446
 Db 440 LysIleValGlySerProAspLeuLysProGluThrSerGluSerIleProGluLeuGlyLeu 459
 QY 1447 TATTACGATAAACCC-----GCCGGTCTGAATGCCAATGTCAAGTTTATG 1494
 Db 460 Ty-TyrMetGlyGluGlyTyLeuGluGlyValGluSerSerValThrValPheArg 479
 QY 1495 ACTGACTTCTCCAAACAGATTGTCTTATTCATAAATGATAACACCAAT----- 1545
 Db 480 AsnAspValLysAspArgIle---SerIleSerArgThrSerAspValAsnAlaAlaPro 498
 QY 1545 ----- 1545
 Db 499 GlyTyThrGlnAsnPheValGlyPheGluThrGlyAlaAsnGlyArgArgIleProValPhe 518
 QY 1546 AGTATGTAAACAGCGGAAAGCCCGGTTCACCGTGTGGAATTTCCCGGCACATGCGC 1605
 Db 519 SerTyTyTrAsnValAsnLysAlaArgAsnGlnGlyValGluThrGluLeuLysIlePro 538
 QY 1606 CTGTGGTTCAGAGGATGTCAGCTGTCACTGAATTACACCTGGACCCCAAGTGAACAACGT 1665
 Db 539 ---PheAsnAspGluTrpLysLeuSerIleAsnTyThrTy-----AsnAspGlyArg 555
 QY 1666 GAT-----CGTGAATAACAAAGTGGCGCTGAGTTATACCCCTGAACACATGGTG 1716
 Db 556 AspValSerAsnGlyGluAsnLys-----ProLeuSerAspLeuProPheHisLeuAla 573
 QY 1717 AATGCGAAACTGAACTGGCAGATCACCGAAGAGTGGCATCATGGCTGGTCCCGTTAT 1776
 Db 574 -----LeuGluAspTrpSerPheTyTrValSerGlyHisTy 585
 QY 1777 CGCGGGAACACACCGTTTCCACCAAGATTATTCCTCAGCGCTCTACAGAGAAA 1836
 Db 586 ThrGly-----GlnLysArg 590
 QY 1837 GTGTATGATGAGAAAGAGAAATACCTGAAAGCTGCGGTGGTGGATCGCTGTCTG 1896
 Db 591 AlaAspSerAlaThrAlaLysThrProGlyGlyTyThrIleTrpAsnThrGlyAlaAla 610
 QY 1897 TGAAGATGACGAGTCCCTGACCTGAGTGTGCGTGAATAACCTGCTCAACAGGAT 1956
 Db 611 TrpGlnValThrLysAspValLysLeuArgAlaGlyValLeuAsnLeuGlyAspLysThr 630
 QY 1957 TACAGTACGCTGAGCTGTACAGTCCGTAAGTACCTGTATCCCGTATCTACTTC 2016
 Db 631 AlaAsnGlyThrLeuAspTrp-----LysProAspLeuSerArgAspAspTy 646
 QY 2017 CAGACGGGATCATCAACACAGGATATGTGATACCTGACGGAATTTACCTGGATGTCCGTG 2076
 Db 647 -----SerTyAsnGluAspGlyArgArgTyPheMetAlaVal 659
 QY 2077 AACTATCAGTTC 2088
 Db 660 AspTyArgPhe 663

RESULT 15

US-09-489-039A-8455
 ; Sequence 8455, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; PRIOR FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 8455
 ; LENGTH: 778
 ; TYPE: PR1

	: ::	:::	:	::		
666	rTgPInValArgLusPluSerLeuGlnSerThrPheThrTrpTy-Gly-----	683				
	::	TCCGGGAAACACCACAGTTTCACCCAGAAATTATCGTCACCTGAGCGGTGCACAGAAGAA	1835			
1776	QY					
		-ArgGlnLysProLysArgPhe-----AsnTyrLysGly--GlulaValSerClySe	700			
684	Db					
		AGTGTAATGATGAAGAGGAGAAATCACTGAAAGCCTGGACGGTGGTGGATGCAGGTCGTGC	1895			
1836	QY					
	::	rgLuAuSnGLuValSerProtyr-----SerIleValcLyLeuSerAlath	716			
700	Db					
		GTCGAAGATGACCGGATGCCCTGACGCTGAATGCTGCGGTGAATAAACCTGCTCAACAAGA	1955			
1896	QY					
	::	rTpAspValAsnLysAsnLeuSerPheThrSerGlyIleAspAsnLeuPhe-----	733			
716	Db					
		TTACAGTAGCGTGACGCTGTACAGTCCCGGTAGAGTACGCTGTATGCCGGTGATTACTT	2015			
1956	QY					
	::	-----AspIleArgHisTyrArgAlaGlyAsnAlaGln-----	744			
734	Db					
		CCAGACGGGATCATCAACACAGA-----	2045			
2016	QY					
		-ThrThrGlyAsnAlaThrThrGlyAlaTyrLeuTyrGlyAlaGlyAlaGluThrTyrAs	764			
745	Db					
		GATACCTGACGCGAAATTACTGGATGTCGCTGAACATACAGTTC	2088			
2046	QY					
		nGuSerGlyValThrPhePheMetSerValAsnThrHisPhe	778			
764	Db					

Search completed: October 13, 2004, 11:56:51
Job time : 71.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: October 13, 2004, 11:31:01 Search time 125 Seconds

(without alignments)

9452.909 Million cell updates/sec

Title: US-10-625-972-4

Perfect score: 3732
Sequence: 1 atcggaataaccactctggc.....cgtgaactatcagttctga 2091

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgn2.1/USPRO_pool/US10625972/runat_12102004_130105_26800/app_query.fasta_1.2247
-DB=A Geneseq 29Jan04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTENT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10625972_QCGN_1_158 -runat_12102004_130105_26800 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Geneseq 29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3632	97.3	696	7 ADC00149	Adc00149 Enterohae
2	3628	97.2	696	2 AAR75366	Aar75366 Adhesin.
3	1029	27.6	660	6 ABM67028	Abm67028 Photorhab
4	972	26.0	682	4 ABB52836	Abbs52836 Escherich
5	617.5	16.5	761	6 ADA34655	Ada34655 Acinetoba
6	604.5	16.2	725	4 AAY72914	Aay72914 E. coli 1
7	539.5	14.5	620	6 ABM68636	Abm68636 Photorhab
8	428.5	11.5	687	6 ABU20598	Abu20598 Protein e
9	399	10.7	684	6 ABU20991	Abu20991 Protein e
10	397.5	10.7	708	2 AAY34501	Aay34501 Porphorym

11	397.5	10.7	772	2 AAY34374	Aay34374 Porphorym
12	391	10.5	660	7 ADC01296	Adc01296 Enterohae
13	388	10.4	660	4 ABB52669	Abbs52669 Escherich
14	383.5	10.3	643	6 ADA33859	Ada33859 Acinetoba
15	373.5	10.0	674	6 ABM67198	Abm67198 Photorhab
16	371	9.9	698	6 ABU49701	Abu49701 Protein e
17	365.5	9.8	663	4 ABB52462	Abbs52462 Escherich
18	352.5	9.4	713	7 ADC01362	Adc01362 Enterohae
19	347.5	9.3	654	4 ABB52577	Abbs52577 Escherich
20	340	9.1	704	6 ADA34250	Ada34250 Acinetoba
21	339	9.1	673	6 ABM67345	Abm67345 Photorhab
22	337.5	9.0	757	2 AAY34472	Aay34472 Porphorym
23	337.5	9.0	763	2 AAY34348	Aay34348 Porphorym
24	327.5	8.8	153	2 AAR60561	Aar60561 Vibrio ch
25	325.5	8.7	713	6 ABU49654	Abu49654 Protein e
26	321	8.6	775	6 ABU20485	Abu20485 Protein e
27	315	8.4	743	6 ADA33786	Ada33786 Acinetoba
28	314	8.4	428	6 ABP57973	Abp57973 Ferric-si
29	314	8.4	428	6 ABR38876	Abt38876 Surface e
30	313.5	8.4	151	2 AAR25587	Aar25587 Truncated
31	310.5	8.3	790	2 AAR95565	Aar95565 N. mening
32	304	8.1	791	2 AAR95566	Aar95566 N. mening
33	302.5	8.1	774	6 ABU28129	Abu28129 Protein e
34	302	8.1	729	6 ABU47292	Abu47292 Protein e
35	301	8.1	763	6 ABU20715	Abu20715 Protein e
36	300.5	8.1	774	4 AAU29336	Aau29336 Novel mar
37	300.5	8.1	774	6 ABU15308	Abu15308 Protein e
38	297.5	8.0	779	6 ABM68355	Abm68355 Photorhab
39	297	8.0	790	2 AAR95568	Aar95568 N. gonorr
40	296.5	7.9	722	3 AAB07697	Aab07697 A. Neisser
41	295	7.9	722	6 ABU27630	Abu27630 Protein e
42	294.5	7.9	766	6 ADA36943	Ada36943 Acinetoba
43	294	7.9	833	2 AAY34494	Aay34494 Porphorym
44	294	7.9	876	2 AAY34368	Aay34368 Porphorym
45	293.5	7.9	719	2 AAW01462	Aaw01462 NTHI Hxuc

ALIGNMENTS

RESULT 1
ADC00149
ID ADC00149 standard; protein; 696 AA.
XX AC
ADC00149;
AC

XX DT 04-DEC-2003 (first entry)

XX DE Enterohaemorrhagic E. coli 0157:H7-specific protein SEQ ID NO: 194.

XX KW enterohaemorrhagic; anti-bacterial.

XX OS Escherichia coli; 0157:H7.

XX PN JP2002355074-A.

XX PD 10-DEC-2002.

XX PF 24-JAN-2002; 2002JP-00015959.

XX PR 24-JAN-2001; 2001JP-00112010.

XX PA (UYTS-) UNIV TSUKUBA.

XX WPI; 2003-451640/43.

XX DR Enterohaemorrhagic Escherichia coli 0157:H7-specific nucleic acid molecule and a polypeptide and its use, a polypeptide, a vector and a host cell.

XX PT Claim 3; SEQ ID NO 194; 2067bp; Japanese.

XX PS The invention relates to a novel enterohaemorrhagic Escherichia coli

XX CC 0157:H7-specific nucleic acid molecule. A polynucleotide of the invention has anti-bacterial activity. The polypeptide can be used in detection

CC and/or treatment of O157:H7 infection. The nucleotide sequence of the
CC genome of Enterohaemorrhagic E coli O157:H7 was determined. The present
CC sequence represents an E. coli O157:H7-specific polypeptide of the
CC invention.

XX
SQ Sequence 696 AA;

Alignment Scores:			
Pred. No.:	0	Length:	696
Score:	.3632.00	Matches:	695
Percent Similarity:	99.86%	Conservative:	0
Best Local Similarity:	99.86%	Mismatches:	1
Query Match:	97.32%	Indels:	0
DB:	7	Gaps:	0

US-10-625-972-4 (1-2091) x ADC00149 (1-696)

Qy	1	ATCGCAATAACCACTCTCGTCTCCGTAGTCAATCCCTCTCGATTTTCAGCCAGCAGC	60
Db	1	MetArgIleThrThrLeuAlaSerValIleProCysLeuGlyPheSerAlaSer	20
Qy	61	ATAGCTGTCGACAGGATCTGATGATTCTCTCGCATCCGGCTATCAGAAAAGCTGACT	120
Db	21	IleAlaAlaIaGluAspValMetIleValSerAlaSerGlyTyrGluIysLeuThr	40
Qy	121	AACGACGCCCGCAGTGTTCCTGATTAGCCAGGAGAAATTGACGTCCAGCCAGTACCAC	180
Db	41	AsnAlaAlaAlaSerValSerValIleSerGlnGluLeuGlnSerSerGlnTyrHis	60
Qy	181	GATCTCGCGGAGCTCTGAGATCAGTAGAGGCTGTGATGTTGAAAGTGGTACGGGTAA	240
Db	61	AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys	80
Qy	241	ACCGAGGGCTCGAAATCAGCATCCGAGGAATGCCAGCCAGTTACAGCGTGATACTGATT	300
Db	81	ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyrThrLeuIleLeuIle	100
Qy	301	GATGTGTTCTGTCAGCGCGGAAGCAGTGACTGCCAACCGTTTTTCTGCCATGAAT	360
Db	101	AspGlyValArgGlnGlyGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn	120
Qy	361	ACCGGTTTCATCCCTCTCGCCCATTCAGAGGTATTCAGGTATTCAGGGGGCCGATG	420
Db	121	ThrGlyPheMetProProLeuAlaAlaIleGluArgIleGluValIleArgGlyProMet	140
Qy	421	TCCACACTATGTGGTCTGTATGCGATGGCGCGTGTGTGAATATCATACCAGAAGAAT	480
Db	141	SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAsnIleIleThrArgIysAsn	160
Qy	481	GNAGACAAATGGCTCTTCCGTCAATCAGCGCTGAATCTGCAGGNAGACACAATCG	540
Db	161	AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp	180
Qy	541	GGTAAACAGCAGCCAGTTTAATTCTCGAGCAGTGGTCCCTCTGTGGATGATTCGTCCAG	600
Db	181	GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspSerValSer	200
Qy	601	CTGACGGTACGCGGTAGACACAACACAGCGTCAGGGTTCATCGGTCAATCATGAGCGAT	660
Db	201	LeuGlnValArgGlySerThrGlnGlnArgGlnGlySerSerValThrSerLeuSerAsp	220
Qy	661	ACAGCAGGCACCGGTATTCCTTATCCACGAGGTACAGAAATATATCTTGTGTGCACT	720
Db	221	ThrAlaAlaThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg	240
Qy	721	CTTGACTCGGAAGCGCTCGGACGAGATGTCTCTGTTTGTATPGGATACCCACCGCGAG	780
Db	241	LeuAspTrpLysAlaSerGluGlnAspValLeuTrpPheAspMetAspThrThrArgGln	260
Qy	781	CGTTATGATAACCGGATGGCACTGGGAGTCTGACGGGGGATATGACCGGACCTTG	840
Db	261	ArgTyrAspAsnArgAspCysGlnLeuGlySerLeuThrGlyGlyTyrAsnArgThrLeu	280


```

Db 641 LeuAsnAlaValAsnAsnLeuLeuAsnLysAspTyrSerAspValSerLeuTyrSer 660
QY 1981 GCCGTAAGACTACGCTGTATGCGGTGATTAATCCAGACGGGATCATCAACAACAGGA 2040
Db 661 AlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThrThrGly 680
QY 2041 TATGTGATACCTGAGCGAAATTAATGCTGATGCTGCTGAACATCAGTTC 2088
Db 681 TyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe 696

RESULT 2
AAR75366
ID AAR75366 standard; protein; 696 AA.
XX AC AAR75366;
XX DT 13-MAY-1996 (first entry)
XX DE Adhesin.
XX KW adhesin; plasmid pear; vector; vaccine; intestine colonisation.
XX OS Escherichia coli.
XX Key Location/Qualifiers
FH Misc-difference 293
FT /note= "deduced residue from nucleotide sequence is His,
FT this differs from Seq ID5, residue is Gly"
XX
XX MO9600233-A1.
XX PD 04-JAN-1996.
XX PF 07-JUN-1995; 95WO-US006994.
XX PR 24-JUN-1994; 94US-00265714.
XX PA (CHIL-) CHILDREN'S HOSPITAL & MEDICAL CENT.
XX PA (UNIW ) UNIV WASHINGTON.
XX PA (UNIW ) UNIV WASHINGTON STATE RES FOUND.
XX PI Tarr PI, Bilge SS, Besser TE, Vary JC;
XX WP; 1996-068826/07.
XX DR N-PSDB; AAT10105.
XX
XX Chromosomal DNA from E. coli O157:H7 encoding epithelial adhesin -
XX isolated on plasmid pSC (overlap), for use as a vaccine to mediate
XX bacterial colonisation of bovine intestine.
XX
XX Disclosure; Page 31-32; 42pp; English.
XX
XX A candidate adhesin (AAR75366) was identified that is a homologue of the
XX IgA protein of Vibrio cholerae. The adhesin enables Escherichia coli
XX O157:H7, an antibiotic-resistant, virulent and common food-borne
XX pathogen, to adhere to epithelial cells. Recombinant adhesin was obtd. by
XX expression of the encoding sequence (see AAT10105) in E. coli HB101
XX (pear). The adhesin can be used as a vaccine for immunisation of cattle
XX against disease or colonisation of mucosal surfaces by O157:H7, thus
XX increasing the safety of food derived from cattle
XX
XX Sequence 696 AA;
SQ

Alignment Scores:
Pred. No.: 0 Length: 696
Score: 3628.00 Matches: 695
Percent Similarity: 99.86% Conservatives: 0
Best Local Similarity: 99.86% Mismatches: 1
Query Match: 97.21% Indels: 0
DE: 2 Gaps: 0

US-10-625-972-4 (1-2091) x AAR75366 (1-696)

```

```

QY 1 ATCGCAATACACACTCTGGCTTCCTAGTATCCCTCTCTCGGATTTTCACCGACGAC 60
Db 1 MetArgIleThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSerSer 20
QY 61 ATAGTGTGTCAGAGGATGTGATGATGTCTCGGCATCCGGCTATGAGAAAAGTGACT 120
Db 21 IleAlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLeuThr 40
QY 121 AACGAGCGCCGAGTGTCTCTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCAC 180
Db 41 AsnAlaAlaAlaSerValSerValIleSerGlnGluLeuGlnSerSerGlnTyrHis 60
QY 181 GATCTGGCGGAGGCTCTGAGATCAGTACAGGCTGTGGATGTTGAAAGTGGTACGGCTAA 240
Db 61 AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys 80
QY 241 ACCGAGGCGCTGGAATCAGCATCCGAGGAATCCAGCAGTTCACGCTGATATCTGATT 300
Db 81 ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyrThrLeuIleLeuIle 100
QY 301 GATGTGTTCTGTCAGGCGGAGCAGTACGCTGCTCCCAACGGTCTTTCTGCCATGAAT 360
Db 101 AspGlyValArgGlnGlyGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
QY 361 ACCGGGTTTCATCCCTCTGCGCCGCTATTGAGCGTATTGAGGTTATCAGGGCGCCGATG 420
Db 121 ThrGlyPheMetProProLeuAlaAlaIleGluAlaGluValIleArgGlyProMet 140
QY 421 TCCACACTGTATGCTCTGATCGGATGGCGGTGGTGGATATCATCACCAGAAGAT 480
Db 141 SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAsnIleIleThrArgLysAsn 160
QY 481 GCAGACAAATGGCTCTCTTCCTCAATGACGCGCTGAATCTGCAGGAAACAAATGG 540
Db 161 AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp 180
QY 541 GGTAAACAGCAGCCAGTTTAATTTCTGAGCAGTGTGCTCCCTTGTGGATGATCTCTCAGC 600
Db 181 GlyAsnSerSerGlnPheAsnPheTrpSerGlyProLeuValAspAspSerValSer 200
QY 601 CTGCAAGTACGGGTAGCACACACAGCGTCAAGGTTTCATCGTCCACATCAGTACGACAT 660
Db 201 LeuGlnValArgGlySerThrGlnGlnArgGlnGlySerSerValThrSerLeuSerAsp 220
QY 661 ACAGCAGCAGCGGTATTCCTTATCCCGAGGATCACAGAATTATAATCTTGGTCCAGT 720
Db 221 ThrAlaGlyThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg 240
QY 721 CTGTGACTGGAAGCGCTCGGACGAGATGTCTCGTTTGATGTATGATACCAACCGCGGAG 780
Db 241 LeuAspTrpLysAlaSerGluGlnAspValLeuTrpPheAspMetAspThrThrArgGln 260
QY 781 CGTTATGATAACCGGGATGGCACTGGGGAGTCTGACGGGGGATATGACCGGACCCCTG 840
Db 261 ArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGlyTyrAspArgThrLeu 280
QY 841 CCCTATGACGGAACAAATTCACGTGCTGATGATCATCTTCCACCTCGGAACATGG 900
Db 281 ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspGlyThrPheThrPheGlyThrTrp 300
QY 901 AAATCGTATCTCAACTGGAACGAGACAGAAATAAAGTCTGAGCTTGTACGCAAGTGA 960
Db 301 LysSerTyrLeuAsnTrpAsnGlnThrGluAsnLysGlyArgGluLeuValArgSerVal 320
QY 961 CTGAACGCGCAAAATGGGGCTTCGCGGTACGCGCGGAGCTTAAGAAATCGAACCTTT 1020
Db 321 LeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu 340
QY 1021 ATCCTGAATTCATTACTCTCTACCCCTCTGGAGATCTCATCTGTTACGCTGGGGGCG 1080
Db 341 IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly 360

```


1081	QY	GAGTTTCAGAGCTCGTCCATGAAGAACCGAGGTGTCTTGTCCAGACAGGTGAACACTTTC	1144
361	Db	GluPheGlnSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe	380
1141	QY	CGGCAGAAAAGCTGGTCCGTATTTCGTGAGGATAGTGGCATCTCACGGATGACATTGCG	1200
381	Db	ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspAlaLeuAla	400
1201	QY	CTGACTCGGGCGACCGCTATGAACATCATAGCAATTCGGGGACCACTTTCAGTCCGCGT	1260
401	Db	LeuThrAlaGlySerArgTyrGluHisGlnPheGlyGlyHisPheSerProArg	420
1261	QY	GCATATCTGCTCGGATGTGGCAGATGCTCTGGACGCTGAAAGCGGTGTGACCAACGGGA	1320
421	Db	AlaTyrLeuValTrpAspValAlaAspAlaTrpThrLeuLysGlyGlyValThrThrGly	440
1321	QY	TATAAGCACCCAGATGGGGCAGCTACATAAAGGATTAGTGTGTCTCCGGCGACGGGA	1380
441	Db	TyrLysAlaProArgMetGlyGlnLeuHisLysGlyLleSerGlyValSerGlyGlnGly	460
1381	QY	AAAAAATCTACTTGTGAACCCCGACCTGAAGACCGGAGAGAGCGTCAGTTATGAGCCT	1440
461	Db	LysThrAsnLeuLeuGlyAsnProAspLeuLysProGluGluSerValSerTyrGluAla	480
1441	QY	GGGTGTATTACGATACCCCGCGCTGGAATGCCAATGTCAAGTTTATGACTGTAC	1500
481	Db	GlyValTyrTyrAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp	500
1501	QY	TTCTCCAAAGATTGCTCTTATCCATAAATGATAACAACAATAGCTATGTAACACG	1560
501	Db	PheSerAsnLysIleValSerTyrSerLleAsnAspAsnThrAsnSerTyrValAsnSer	520
1561	QY	CGAAGCCCGGTTCACGGTGGGAATTGCGCGCACATTGCCGCTGTGTGTACAGAGAT	1620
521	Db	GlyLysAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp	540
1621	QY	GTCACGCTGTCACTGAATTACACTCGGACCCGGAAGTGAACAAACGTGATGTATAACAA	1680
541	Db	ValThrLeuSerLeuAsnTyrThrTrpThrArgSerGluGlnArgAspGlyAspAsnLys	560
1681	QY	GGTCCGCGCTGAGTTATACCCCTGAACACATGCTGAATGCGAACTGAATCGGCAGATC	1740
561	Db	GlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle	580
1741	QY	ACCAAGAGGTGGCATCATGCTCGGTGCGCTTATCGCGGAGAAACACCACTTTCACC	1800
581	Db	ThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgGlyLysThrProArgPheThr	600
1801	QY	CAGAATTATTCTGCTACGTACCGCTGTACAGAGAAGTGTATGATGAGAAGGAGATAC	1860
601	Db	GlnAsnTyrSerSerLeuSerAlaValGlnLysLysValTyrAspGluLysGlyGluTyr	620
1861	QY	CTGAAGCCTGGACGGTGGTGGATGACAGTCTGTCTGTGGAAGATCACGGATGCCCTGACG	1920
621	Db	LeuLysAlaTrpThrValValAspAlaGlyLeuSerTrpLysMetThrAspAlaLeuThr	640
1921	QY	CTGAATCTCGGTGTAATAACTGCTCAACAAGGATTACGTGACGTGAGCCTGTACAGT	1980
641	Db	LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyrSerAspValSerLeuTyrSer	660
1981	QY	GCCGCTAAGAGTAGCTCGTGTATGCGGTGTATTCTCCAGACGGGATCATCAACAACAGGA	2040
661	Db	AlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThrThrGly	680
2041	QY	TATGTGATACCTGACGGAAATTACTGGATGTCGCTGAACATTCAGTTC	2088
681	Db	TyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe	696
RESULT 3			
ABM67028			
ID	ABM67028	standard; 660 AA.	
XX			
AC	ABM67028;		

AC ABB52836;
 XX
 DT 11-FEB-2002 (first entry)
 XX
 DE Escherichia coli polypeptide SEQ ID NO 1077.
 XX
 KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
 XX immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
 XX systemic infection; non-diarrhoeal infection; septicaemia;
 XX pyelonephritis; antibiotic resistance.
 XX
 OS Escherichia coli.
 XX
 XX WO200166572-A2.
 PN
 XX 13-SEP-2001.
 PD
 XX 12-MAR-2001; 2001WO-EP003445.
 XX
 XX 10-MAR-2000; 2000FR-00003145.
 PR
 XX 02-FEB-2001; 2001FR-00001449.
 XX
 XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA
 XX Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
 PI
 XX WPI; 2001-550253/61.
 XX
 XX A library of DNA fragments of Escherichia coli strains for the phylogenetic
 PT determination of a given strain comprises polynucleotides of nature B2/D+
 PT A-.
 XX
 XX Example 6; Fig 6; 646pp; English.
 XX
 XX The invention relates to a library of DNA fragments of Escherichia coli
 CC strains comprising polynucleotides (AB48577-AB486729 and AB489533) and
 CC encoded proteins (AB52459-AB52919 and AB52954-AB53094) of nature
 CC B2/D+A-. The polynucleotides have potential antiinflammatory,
 CC antibacterial and immunosuppressive activity as part of pharmaceutical
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli
 CC infections. The polypeptides are useful for determining the phylogenetic
 CC group of a given E. coli strain. These polypeptides can detect and treat
 CC an undesired development of E. coli, particularly an extra-intestinal
 CC infection that include systemic and non-diarrhoeal infections such as
 CC septicaemia, pyelonephritis and meningitis this is particularly
 CC advantageous as bacterial resistance is increasing with the more frequent
 CC use of broad spectrum antibiotics
 XX
 XX Sequence 682 AA;
 SQ

Alignment Scores:
 Pred. No.: 2,896-81 Length: 682
 Score: 972.00 Matches: 244
 Percent Similarity: 51.08% Conservative: 135
 Best Local Similarity: 32.88% Mismatches: 245
 Query Match: 26.05% Indels: 118
 DB: 4 Gaps: 23

US-10-625-972-4 (1-2091) x ABB52836 (1-682)

QY 7 ATACCACTCTCGCTTCGATGATTCCTCTCGGATTTTCAGCCAGCAGCATAGCT 66
 DB 11 llaAlaValMetCysSerAlaValIleSerSer---GlyTyrAlaSerSerAsp----- 27
 QY 67 GCTGCAGAGGATGTGATGATTCCTCGGATTCGCGATCCGCTATGAGAAAGCTGACTAACGCA 126
 DB 28 LysLysGluAspThrLeuValThrAlaSerGlyPheThrGlnGlnLeuArgAsnAla 47
 QY 127 GCCGCCAGGTTTCTGTGATTCAGCAGGAGGATTCAGCCAGTCAGTCAGCCAGTACCATCTG 186
 DB 48 ProAlaSerValSerValIleThrSerGluGlnLeuLysLysProValSerAspLeu 67
 QY 187 GCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAGTGGTACGGGTAAACCCGGA 246

DB 68 ValAspAlaValLysAspValGluGlyIleSerIleThrGlyGlyAsnGluLysPro--- 86
 QY 247 GGGCTGGAATCAGCATCCGAGGAATGCCAGCAGTATACCGCTGATCTGATGATGATGAT 306
 DB 87 -----AspIleSerIleArgGlyLeuSerGlyAspTyrThrLeuIleLeuValAspGly 104
 QY 307 GTTCGTGAGGGGGAAGCAGTACGCTGACTCCCAACGGTTTTCGCCATGAATACCCGG 366
 DB 105 ArgArgGlnSerGly---ArgGluSerArgProAsnGlySerGlyGlyPheGluAlaGly 123
 QY 367 TTCATGCCCTCTCGCCGCCATTCAGCGCTATGTAGGTTATCAGGGGCCGCGATGCCACA 426
 DB 124 PheIleProValGluAlaIleGluArgIleGluValIleArgGlyProMetSerSer 143
 QY 427 CTGTATGGCTCTGATCGCATGGCGGTGTGGTGAATATCATATACCATCAGAAAG---AATGCA 483
 DB 144 LeuTyrGlySerAspAlaIleGlyValIleAsnIleIleThrLysProValAsnAsn 163
 QY 484 GACAAATGGCTCTCTCCGCTCAATGCAGGGCTGAATCTGCAGAAAGCAACAAATGGGGT 543
 DB 164 GlnThrTyrAspGlyValLeuGlyLeuGlyIleIleGlnGluHisGlyLysPheGly 183
 QY 544 AACAGCAGCCAGTTAAATTTCTGGAGCAGTGTCCCTTGTGGATGATCTCTCAGCCCTG 603
 DB 184 AsnSerThrThrAsnAspPheTyrLeuSerGlyProLeuIleLysAspLysLeuGlyLeu 203
 QY 604 CAGGTACGGGTAGCAGACACACAGCGCTCAGGGTTCATCGTCCATCATCTAGCGCATACA 663
 DB 204 GlnLeuTyrGlyGlyMetAsnTyrArgLysGluAspSerIleSerGln----- 219
 QY 664 GCAGGCAGCGCTATTCCTTATCCACGGAGTCCAGAAATTAATATCTTGGTCAGCTGTT 723
 DB 220 ---GlyThr-----ProAlaLysAspAsnLysAsnIleThrAlaThrLeu 233
 QY 724 GACTGGAAGCGCTCGGAGCAGGATGCTCTGCTTTGATATGATATACACCCGCGAGCGT 783
 DB 234 GlnPheThrProThrGluSerGlnLysPheValPheGluTyr----- 247
 QY 784 TATGATAACCGGATGGGCAACTCTGACGGGGGATATGACCGGACCTGCGC 843
 DB 248 -----GlyLysAsnAsnGlnValHisThrLeuThrProGlyGluSerLeuAspAlaTrp 265
 QY 844 TATGACGGAACAAATTTACGTGCTGCTATGATATCTTACCTTCGGAACATGGAA 903
 DB 266 ThrMetArgGlyAsnLeuLysGlnProAsnSerLysArgGluThrHisAsnSerArgSer 285
 QY 904 TCGTATCTG---AACTGGAAAC-----GAGACAGAAAAATAA 936
 DB 286 HisTrpValAlaAlaTrpAsnAlaGlnGlyGluIleLeuHisProGluIleAlaValTyr 305
 QY 937 GGTCGTGAGCTGTACGCGAGTGTACTG-----AGCGCGACAAA-----TGGGG 981
 DB 306 GlnGluLysValIleArgGluValLysSerGlyLysLysAspLysTyrAsnHisTrpAsp 325
 QY 982 CTTCGCGCTCAGCGCGAGCTTAAGGAATCGAATCTTATCTGAAATTCATTACTGCTT 1041
 DB 326 LeuAsnTyrGluSerArgLysProGluIleThrAsnThrIleLeuAspAlaLysValThr 345
 QY 1042 ACCCTCTCGGAGAAATCTCATCTGTTACGGTGGGGGCGAGTTTCAGAGCTGCTGCATG 1101
 DB 346 AlaPheLeuProGlu---AsnValLeuThrIleGlyGlyGlnPheGlnHisAlaGluLeu 364
 QY 1102 AAGACGGAGTTCTCTCCCTCCAGCAGCGTGA-----ACT 1137
 DB 365 ArgAspAsp-----SerAlaThrGlyLysLysThrThrGluThrGlnSerValSer 381
 QY 1138 TTCGCGCAGAAAAGCTGTCGCTATTTCTGAGGATGATGATGATCTCTCAGGATCATT 1197
 DB 382 IleLysGlnLys-----AlaValPheIleGluAsnGluTyrAlaAlaThrAspSerLeu 399
 QY 1198 GCGCTACTCGCGGAGCGCTATGACATCATGACATTCGCGGGGACACTTCAGTCG 1257

XX	AAV72914;
AC	
XX	
DT	13-JUN-2001 (first entry)
XX	
DE	E. coli ironNec extracytoplasmic protein fragment.
XX	
KW	ironNec; extracytoplasmic protein; immunogen; vaccine; ET; UTI;
KW	immunotherapy; extraintestinal infection; urinary tract infection;
KW	meningitis; pneumonia; intra-abdominal infection; antibiotic.
XX	
OS	Escherichia coli.
XX	
FH	Key Location/Qualifiers
FT	Peptide l..24
FT	/label= Signal_peptide
FT	Protein 25..725
FT	/label= Mature_E. coli_ironNec_extracytoplasmic_protein
FT	Misc-difference 120..121
FT	/note= "Encoded by GTG CGT TAT"
FT	Misc-difference 134..136
FT	/note= "Encoded by AAC TGG"
XX	
PN	WO200121636-A1.
XX	
PD	29-MAR-2001.
XX	
PF	22-SEP-2000; 2000WO-US026117.
XX	
PR	22-SEP-1999; 99US-0155621P.
XX	
PA	(UUYNY) UNIV NEW YORK STATE RES FOUND.
XX	
PI	Russo T, Carlino U;
XX	
DR	WPI; 2001-244936/25.
XX	
N-	PSDB; AAD03054.
XX	
PT	Novel isolated ironNec polynucleotide from extraintestinal isolate of
PT	Escherichia coli useful as vaccine for treating or preventing
PT	extraintestinal infections caused by extraintestinal pathogenic
PT	Escherichia coli.
XX	
XX	Claim 9; Page 39-41; 44pp; English.
XX	
CC	The invention relates to the identification of ironNec gene, from an
CC	extraintestinal isolate of Escherichia coli. This gene is expressed in
CC	increased amounts in human urine and is identified by transposon (TnphoA)
CC	mutagenesis. ironNec gene encodes an extracytoplasmic protein. This gene
CC	can be used as an immunogen in vaccine formulations. The recombinant
CC	vector comprising nucleotide sequence encoding one or more antigenic
CC	epitope of ironNec is useful for diagnostic and immunotherapeutic
CC	purposes. The ironNec antigenic peptide is useful for treating or
CC	preventing extraintestinal infections (EIs) caused by extra-intestinal
CC	pathogenic E. coli (EXPEC). The EIs include urinary tract infection
CC	(UTI), meningitis, intra-abdominal infection and pneumonia. The present
CC	sequence is E. coli ironNec extracytoplasmic protein
XX	
SQ	Sequence 725 AA;
XX	
Alignment Scores:	
Pred. No.:	6.3e-47 Length: 725
Score:	604.50 Matches: 212
Percent Similarity:	43.88% Conservative: 136
Best Local Similarity:	26.73% Mismatches: 280
Query Match:	16.20% Indels: 165
DB:	Gaps: 34
US-10-625-972-4 (1-2091) x AAV72914 (1-725)	
Oy	1 ATCGGAATAACC---ACTCTGGCTTCGTAGTCATCCTGTCGGATTTCAGGCAGC 57
Dd	1 MetArgilleAsnlystlelueTrpSerLeuThrValLeuVaIGlyLeuAsn---Ser 111

QY 1927 GCTGGCGTGAATACCTGCTCAACAGGATTAACAGTACGCTGAGCGCTTACAGTCCGGT 1986
Db 588 AlaArgGileAlaAsnLeuLeuAspLysAspTyr----- 598
QY 1987 AAGAGTACGCTGATGCGGTGATTAATCTCCAGACGGGATCATCAACACAGGATATGTG 2046
Db 599 ---GluThrValTyr-----GlyTyrArg 605
QY 2047 ATACCTGAGCGAATTAATCTGGATGTCGCTGAATATCACTTC 2088
Db 606 IleProGlyArgGluTyrTyrLeuThrGlySerTyrAsnPhe 619
RESULT 8
ABU20598
ID ABU20598 standard; protein; 687 AA.
XX AC ABU20598;
XX DT 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #6125.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Bacteroides fragilis.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX DR WPI; 2003-029926/02.
XX DR N-PSDB; ACA24468.

XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 48522; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 687 AA;

Alignment Scores:
Pred. No.: 1.68e-30 Length: 687
Score: 428.50 Matches: 191
Percent Similarity: 40.00% Conservative: 125
Best Local Similarity: 24.18% Mismatches: 261
Query Match: 11.48% Indels: 213
DB: 6 Gaps: 38

US-10-625-972-4 (1-2091) x ABU20598 (1-687)

QY 16 CTGGCTCCGCTAGTCACTCCCTGCTCTCGGATTTTCAGCCAGCAGCATAGCTGTCGAGAG 75
Db 8 LeuLeuLeuThrLeuLeuGlyThrLeuProLeuAlaAlaGlnAsnValArgGlnGluGln 27
QY 76 GATGTG-----ATGATTGCTCGGCA 96
Db 28 AspThrValSerTyrMetAsnAspPropheAsnLeuGluGlnLeuValValThrAla 47
QY 97 TCGGGCTATGAGAAAAGCTGACTACGCGCGCGCGAGTGTCTCTGTTATTAATCCAGGAG 156
Db 48 ThrArgThrGluLysLysLysLysThrProValIleThrGluIleThrSerIys 67
QY 157 GAATTCAGTCCAGCAGCAGTACCCAGATCTGGCGAGGCTCTG---AGATCAGTAGAGGT 213
Db 68 GlnIleGluGluArgGlyThrGlyAsnIleGlnAspLeuLeuThrGlnGluValProGly 87
QY 214 GTGGATGTTGAAAGTGTACGGGTAAACCGCA---GGCTGGAATCAGATCCGAGGA 270
Db 88 LeuAsnPheGln-----GluValGlyTyrGlyThrSerIleAspIleGly 103
QY 271 ATGCCAGCCAGTTACACGCTGATCTGATTTGATGTTGTTCTGT-----CAGGCGC 318
Db 104 LeuGlySerLysHisIleLeuPheLeuIleAspGlyGluArgIleAlaGlyGluAsnGly 123
QY 319 GGAAGCAGTACGCTGACTCCCAACGGTTTTTCTGCCATGAATACCGGGTTTCATGCCCCCT 378
Db 124 GlyAsnIleAsp-----TyrSerArgIleAsn----- 132
QY 379 CTGGCGCCGCTATGAGGTATTGAGTTCATCAGGGCGCGGATGTCACACTGTATGCTCT 438
Db 133 LeuTyrAsnIleAspHisIleGluLeuValLysGlyAlaSerSerAlaLeuTyrGlySer 152
QY 439 GATGCGATGGCGGTGTGTGTAATTCATTACCAAGAAATGCAGACAATGGCTCTCT 498
Db 153 GlnAlaMetGlyGlyValIleAsnIleThrArgLysAlaLysLysLysPhe----- 170
QY 499 TCCGTCATGACGGGCTGAATCTCGCAAGAAACAAATGGGTACAGCAGCCAGTTT 558
Db 170 ----- 170
QY 559 AATTCTTGAGCAGTGTGTCCTTGTGATGATTCGTACGCTGACAGTACGGGTAGC 618
Db 171 -----GluAlaSerAlaGlyIleArgTyrAlaGlyArg 181
QY 619 ACACACACGCTCAGGGTTCATCGGTCTACATCCTGAGGATACAGCAGGACGGTATT 678
Db 182 AsnGlnGln-----AsnTyrLysAspThrProLysAspHisSer 194
QY 679 CTTTAT-----CCACGGAGTCCACAGAAATTAATATCTTGGT 714

Db	195	GlnTyrLysTyrArgIleHisIleuAspLysProAsnLeuAsnThrAsnLeuSerLeuGly	214
Qy	715	GCACGTCCTTCACTGGAGCGCTCGGACGAGATGCTCTCTGTTGATATGATACACACC	774
Db	215	LeuAsnLeuGlyLysPheThrMetAsnThrAspValIleuTyrLysSerPheAsp	233
Qy	775	CGCAGCGTTATCATACCGG	795
Db	234	TyrGlnLeuPheAspLysLysProLeuValLysTyrPheProAlaTyrAsnThrThrIle	253
Qy	796		810
Db	254	ThrGluGluLeuSerLysProProThrSerIleSerGlyTyrGluAspValGlnValAla	273
Qy	811		831
Db	274	HisLysMetAspTyrArgPheSerLysArgLeuLysValGlnLeuLysGlySerTyr	292
Qy	832	CGGACCTCGGCTATGAGCGAAACAAATTCAGCTGGCTATCATCATCTTTC	885
Db	293		308
Qy	886		933
Db	309	SerGluAspTyrThrTyrGlySerIleAspTyr	323
Qy	934	AAAGCTCGTGAGCTTGACCGAGTGA	978
Db	324	AspLysSerSerLeuValAlaSerValHisThrAspHisTyrAsnArgTyrAspLysTyr	343
Qy	979	GGGCTT---GCCGGTCACGCCCGGAGCTTAAAGAAATCGAACTTATCTCTGAATTCATTA	1035
Db	344	GluLeuLysSerGlyArgGluGluTyrLysAsnAsnIleIleGlnProArgIleVal	363
Qy	1036	CTGCTTACCCTCTGGGAGAACTCTCATCTGGTTACGGTGGGGCGAGTTTCAGAGCTCG	1095
Db	364	TyrSerThrThrAlaLeuAspLysGlnThrIleThrGlyLeuGluTyrTyrArgGlu	383
Qy	1096	TCCATGAAAGACGGAGTTGTCTTTCGCAGCACAGGTGAAACTTTCGGCGCAGAAAAGCTGG	1155
Db	384	SerLeuPheSerAspLysPhe	401
Qy	1156		1209
Db	402	TyrAlaThrAlaLeuGlnAspAspTyrSerIleAsnLysGlnPheSerValIleAla	421
Qy	1210	GGCAGCGCTATGAACATCATAGCAATTCGGGGACACTTCAGTCGCGCTGCATATCTG	1269
Db	422	GlyLeuArgCysAspTyrHisGluLysTyrGlyThrAsnLeuThrProLysAlaSerVal	441
Qy	1270	GTCGTGGATGTGGCAGATGCTCGGACGCTGAAAGCGGTGTGACCGGATATAAGCCA	1329
Db	442	MetTyrLysIlePhe	460
Qy	1330	CCAGAAATGGGCGACGTACATAAAGGATTAGTGTGTCCGGCGAGGAAAAACAAT	1389
Db	461	ProSerIleLysGluLeuTyrMetAsnTrpAspHisIleu	477
Qy	1390	CTACTTGGTAACCCCGACCTGAAGCCGGAAGAGAGC	1443
Db	478	IleTyrGlyAsnSerLysLeuLysProGluThrAsnAsnTyrIleSerLeu	496
Qy	1444	GTGTATTACGATAACCCCGCGGTCTGAATGCCAATGTCACAGGTTTTTATGACTGATTC	1503
Db	497	GluTyrValAsnSerTyrIleAsnIleAsnAlaAsnVal	514
Qy	1504	TCCAAACAAGATTGCTCTTATTTCCTAATAATGATACACCAAT	1560
Db	515	ArgAsnLysIleGluGlyMetTrpSerAsnAspGlnThrGluLeuHisTyrIleAsnIle	534
Qy	1561	GGAAAGCCCGGTTCACAGGTGTGGAATTCGCCGCACATTCGCCGCTGTGCTCAGAG	1617
Db	535	GlyLysSerArgLeuAlaGlyValGluThrMetCysLysIleGlnIleAsnArgHisIle	554

Qy	1618	GATGTCACGCTGTCACCTGAATTACACCTCGGACCCGGAAGTGAACAACGCTGATGCTGATAC	1677
Db	555	AsnValHisGlyAlaTyrAsnTyrLeuTyrThr-----SerLysAspAlaAsp----	570
Qy	1678	AAAGVTCGCGCGCTGAGTTATACCCCTGAACACATCGTGAATCGGAACTGAACTGGCAG	1737
Db	571	---GlyValArgLeuSerSerSerProHisSerGlyAsnIleArg-----	585
Qy	1738	ATCACCCGAAAGAGTGCATCATGGCTGGTGCCCGCTTATCGCGGGAAAAACACACCGTTTC	1797
Db	586	-----AlaGluTyrAsnThrArgIleProArgTyr	595
Qy	1798	ACCAG-----AATTATTGCTCCTCAGCGCTGTACAGAAGAAA-----GTGTATGAT	1845
Db	596	AlaThrValValAsnLeuSerGlyAsnIleMetGlyLysLysPheAspValLeuAsp	615
Qy	1846	GAG-----AAAGGAGCAATACCTCAAAGCCTCG-----ACGGTGGTGGATGACGGT	1890
Db	616	GluLeuGluIleAspGlyLysValGluAlaTyrTyrGlnAlaLysValAsnProTyr	635
Qy	1891	CTGTGCTGGGAAGATGACG-----GATGCCCTGACGCTGAATGCTGCG	1932
Db	636	CysLeuTyrAspLeuThrValSerGlnTyrIleMetGlnAsnLeuArgIleThrAlaGly	655
Qy	1933	GTGAATAACTGCTCAACAGGATTACAGTGACGTGAGCCTGTACAGTGCCTGTAAGAGT	1992
Db	656	IleThrAsnLeuPheAsp-----TyrThrSerAspArgVal	667
Qy	1993	ACGTGTATGCCGCTGATTACTTCACAGCGGATCATCAACAACAGGATATGTGATACCT	2052
Db	668	Thr-----PheAsnThrSerThrSer-----Pro	675
RESULT 9			
ID	ABU20991	standard; protein; 684 AA.	
XX	AC	ABU20991;	
XX	AC	ABU20991;	
DT	19-JUN-2003	(first entry)	
XX	DE	Protein encoded by Prokaryotic essential gene #6518.	
XX	DE	Antisense; prokaryotic essential gene; cell proliferation; drug design.	
XX	KW	Bacteroides fragilis.	
XX	OS	Bacteroides fragilis.	
XX	FN	WO200277183-A2.	
XX	PD	03-OCT-2002.	

RESULT 9

ABU20991	
ID	ABU20991 standard; protein; 684 AA.
XX	
AC	ABU20991;
XX	
DT	19-JUN-2003 (first entry)
XX	
DE	Protein encoded by Prokaryotic essential gene #6518.
XX	
KW	Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX	
OS	Bacteroides fragilis.
XX	
PN	WO200277183-A2.
XX	
PD	03-OCT-2002.
XX	
PF	21-MAR-2002; 2002WO-US009107.
XX	
PR	21-MAR-2001; 2001US-00815242.
PR	06-SEP-2001; 2001US-00948993.
PR	25-OCT-2001; 2001US-0342923P.
PR	08-FEB-2002; 2002US-00072851.
PR	06-MAR-2002; 2002US-0362699P.
XX	
PA	(ELIT-) ELITRA PHARM INC.
XX	
PI	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI	wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX	
DR	WPI; 2003-029926/02.
DR	N-PSDB; ACA24861.
XX	
PT	New antisense nucleic acids, useful for identifying proteins or screening
PT	for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 48915; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 684 AA;

Alignment Scores:
 Pred. No.: 9-52e-28 Length: 684
 Score: 399.00 Matches: 164
 Percent Similarity: 42.5% Conservative: 131
 Best Local Similarity: 23.6% Mismatches: 264
 Query Match: 10.6% Indels: 134
 DB: 6 Gaps: 34

US-10-625-972-4 (1-2091) x ABU20991 (1-684)

QY 73 GAGGATGTGATG-----ATTGCTCGGCATCCGGCTATGAGAAAAAGCTG 117
 DB GluAspGlnMetAsnLeuSerThrValValIleThrGlyThrArgThrProLysLeuLeu 91
 QY 118 ACTAACCGCAGCGCCAGTGTCTCTGATTAGCCAGGAGGAATTGTCAGTCCAGCCAGTAC 177
 DB LysAspAlaProIleIleThrArgValIleThrAlaGlyAspLeuLysValAspAla 111
 QY 178 CACGATCTGCGCGAGGCTCTGAGATCAGTACGAGGTGTGGANGTTGAAAGTGTACCGGT 237
 DB ThrHisIleGlyGlnLeuLeuGln---ValGluLeuProGlyIleGluPheSerTy-Ser 130
 QY 238 AAAACCGGAGGCTGGAATCAGATCCGAGGAATGCCAGCAGTTCACCGCTGATACCTG 297
 DB MetAspGlnGlnValLysLeuAsnMetGlnGlyPheGlyGlyAsnAlaValLeuPheLeu 150
 QY 298 ATTGATGCTGTGTCGAGCGGCGAGCAGTACGCTACTCCCAACGGTCTTTCTGCCATG 357
 DB ValAspGlyGluArgLeuAlaGlyGluThr---LeuAspAsnIleAspTyAsnArgLeu 169
 QY 358 AATACCGGTTTCATGCCCTCTGCGCCCATTCAGGCTATGAGGTATCAGGGGGCCG 417
 DB Asn-----LeuAspAsnValGluArgValGluIleValLysGlyAla 183
 QY 418 ATGTCCACACTGTATGGCTCTGATGCGATGGCGGTGTGTGTAATCATATTACCAAG 477

DB AlaSerThrLeuTyGlySerSerAlaIleGlyValIleAsnIleIleThrLysAla 203
 QY 478 AATGCACACAAATGGCTCTCTCCGTCATGTCAGGCTGAATCTGCAGGAAACACACAAA 537
 DB SerAspAspProIlePheAsnLeuAsnLeuAsnThrArgPheGlyValHisAsnAspGlnArg 223
 QY 538 TGGGGTAAACAGCAGCCAGTTTAAAT-----TTCTGGAGCAGTGTGTCCTCCCTTGGAT 588
 DB HisGlyGlyThrValGlyPheAsnAlaGlyLysPheTy-SerGlnThrAsnValGlnTy 243
 QY 589 GATCTGTC---AGCTGTCAGTACGC---GGTAGCACACACACAGCTCAGGTTTCATCG 642
 DB ThrAsnIleAspSerIleHisValLysGlnGlyAspTyThrThrIleAsnGlyAsnLys 263
 QY 643 GTCACATCATCTAGCAGTACATACAGCAGCAGCGCTATTCCTTATCCACGAGTACAGAAAT 702
 DB ThrTrpAsnValLysGlu-----ArgLeuMetPhe---ThrProAsnGluGln 278
 QY 703 TATAATCTTGGTCAGCTCTTGACTGGAAGGCGTCGAGCAGGATGTCTCTGTTGAT 762
 DB LeuArgLeuThrAlaArgAlaGlyTyTyTyPheArgGluArgAspAla----- 294
 QY 763 ATGGATACACACCGCAGCGTATGATAACCGGATGGCAACTGGGAGTCTGACGGGG 822
 DB SerSerGluThrLysAsnArgTy-----ArgGlyPheSerGlyGlyLeuLysGly 311
 QY 823 GGATATGACCGGACCTGCGCTATGAGCGAAACAAATTCAGTGGCTATCATCAT 882
 DB AsnTyAsp-----PheAsnThr---LysSerAsnLeu----- 314
 QY 883 TTCACCTTCGGACATGGAATCGTATCTCACTGGAACAGACAGACAGAAATAAGTCTG 942
 DB -----PheAsnThr---LysSerAsnLeu----- 321
 QY 943 GAGCTTGATCGCAGTGTACTGAAGCGCGACAAATGGGG---CTTCCGCGTCAGCGCGG 999
 DB GluLeuAlaTyThrPheAspGlnTyAspLysSerAspTyLeuValSerTyLysAsn 341
 QY 1000 GAGCTTAAGGAATCGAACCTTATCCTGAATTCA-----TTACTGCTTACCTCTG 1050
 DB AspIleArgAspTySerAsnValGlnHisSerValArgAlaLeuTyAsnTyThrPhe 361
 QY 1051 GGAGAATCTCATCTGTTACGGTGGCGGCGAGTTCAGAGCTCGTCATGAAAGACGA 1110
 DB AsnAspLysAsnThrLeuThrValGlyGlyAspTy-----LeuArgAspTy 377
 QY 1111 GTTGCTCTCCAGCAGCAGGTGAAACT-----TTCCGCGAGAAAGCTGTCGTTATTT 1164
 DB LeuMetSerTyGlnPheLysGluAsnAlaAspTyThrMetHisSerAlaAspAlaPhe 397
 QY 1165 GCTGAGGATGAGTGGCATCTCAGCATGCTGCGTGGCTGACCTGCGGAGCGGCTATCAA 1224
 DB GlyGlnPheAspTyPheAsnProThrGluHisPheAsnValIleAlaGlyLeuArgPheAsp 417
 QY 1225 CATCATGAGCAATTCGGG---GGACACTCAGTCCGCTGCTATCTCTGCTGCGATG 1281
 DB TyrPheSerGluSerAsnValArgHisPheSerProHisLeuGlyLeuMetTyLysIle 437
 QY 1282 CGAGATCCTCGACCGCTGAAAGCGGCTGTGACACCGGATATAGGCCACCCAGATGGG 1341
 DB GlyAsnCys---SerLeuArgGlySerTyAlaGlnGlyPheArgSerProThrLeuLys 456
 QY 1342 CAGCTACATAAAGGATAGTGTGTGTGTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1398
 DB GluMetHisMetAsnPheTy-----MetAlaAsnThrMetMetIleTyGly 472
 QY 1399 ACCCGGACCTCAACCCGAGAGAGCGTCAGTATAG---GCTGGGCTGATATCAGAT 1455
 DB AsnProAspLeuGluProGluThrSerHisAsnPheSerLeuSerGlyGlyTyThrLys 492
 QY 1456 AACCCCGCGCGTCTGAATGCCAATGTCAAGGTTTATGACTGCTTCTCCCAAGATTT 1515

Db 493 Asn-----ArgTyrAsnPheThrLeuThrGlyTyrTyrAsnLeuValHisAspArgIle 510
Qy 1516 GTCTCTTATTCATTAATGATAAC-----ACCAATAGCTATGTAACACGGGA 1563
Db 511 GluTyrThrSerPheArgAspThrAspGlyMetIleAlaGlnLysTyrIleAsnThrPro 530
Qy 1564 AAGCCCGGTGACGGGTGCGAATTTCCCGGCACATTGCGCTGGTCAGAGGATGTC 1523
Db 531 ArgValAspIleAlaGlyIleAsp---AlaAsnAlaSerAlaLysTyrProCysGlyIle 549
Qy 1624 ACCTGTCTCACTGAATTAACCTCGACCGAAGTGAACACAGCTGATGCTGATTAACAAGGT 1683
Db 550 GlyAlaArgIleSerTyrThrTyrIleHisclupHemMetArgAspGlyGlnThrLys--- 568
Qy 1684 GCGCCGTGAGTTATACCTGGAACACATGCTGTAATCGGAACTGAACCTGCGACATCAC 1743
Db 569 -----LeuSerSerThrArgProHisSerAlaThrValArgLeuGluTyr----- 583
Qy 1744 GAAGAGTGGCATCATGCTGGTGGTGGCCGTTATCGCGGAAACA---CCACGTTTCACC 1800
Db 584 -----GlyLysThrTyrAspHisTyrAsp 591
Qy 1801 CAGAAATTATTCG-----TCCTGAGCGCTGTACAGAAAGAGTGTAT----- 1842
Db 592 PheAsnLeuSerLeuAspGlyArgAlaLeuSerGlnValLysThrAsnGlnTyrThrSer 611
Qy 1843 -----GATGAGAAAGGAGAACTACCTGAAAGCTGCGACGGTGGTGGAT 1884
Db 612 AsnAspProAsnAlaGlyThrGluLysValThrTyr---ProGlyTyrThrMetTrpAsn 630
Qy 1885 GCAGGCTCTGCTGGAAGATGACGGATGCGCTGACGCTGAATGCTGCGGTGAATAACCTG 1944
Db 631 LeuThrLeuThrGlnArgValTrpLysGlyIleAsnValAsnMetAlaValAsnLeu 650
Qy 1945 CTCACAAAGGATTACAGTACGCTGACGCTGACAGCTGACAGTCCCGTAAGAGTACGCTGATGCC 2004
Db 651 PheAsn-----TyrArg 654
Qy 2005 GGTGATTACTTCCAGACGGGATCATCA---ACAACAGGA 2040
Db 655 ProAspTyrTyrTyrAlaAsnSerProTyrThrThrGly 667
RESULT 10
ID AAY34501
XX AAY34501 standard; protein; 708 AA.
AC AAY34501;
XX 27-AUG-2003 (revised)
DT 20-MAR-2003 (revised)
DT 25-AUG-1999 (first entry)
XX Porphorymonas gingivalis protein PG40.
XX Porphorymonas gingivalis; PG; periodontal disease; gingivitis; vaccine;
XX antigenic.
XX Porphorymonas gingivalis.
XX WO929870-A1.
XX 17-JUN-1999.
XX 10-DEC-1998; 98WO-AJ001023.
XX 10-DEC-1997; 97AU-00000839.
PR 31-DEC-1997; 97AU-00001182.
PR 30-JAN-1998; 98AU-00001546.
PR 10-MAR-1998; 98AU-00002264.
PR 09-APR-1998; 98AU-00002911.
PR 23-APR-1998; 98AU-00003128.
PR 05-MAY-1998; 98AU-00003338.
PR 22-MAY-1998; 98AU-00003654.

PR 29-JUL-1998; 98AU-00004917.
PR 30-JUL-1998; 98AU-00004963.
PR 04-AUG-1998; 98AU-00005028.
XX (CSLC-) CSL LTD.
XX Ross BC, Barr IG, Patterson MA, Agius CT, Rothel LJ, Margetts MB;
PI Hocking DM, Webb EA;
XX WPI; 1999-385613/32.
DR N-PSDB; AAX91719.
XX Antigenic Porphorymonas gingivalis peptides for preventing gingivitis.
PT Claim 1; Page 482-483; 588pp; English.
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX1802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphorymonas gingivalis. Probes can
CC be used to detect Porphorymonas gingivalis in standard hybridisation
CC assays. Porphorymonas gingivalis is involved in periodontal disease
CC especially gingivitis. (Updated on 20-MAR-2003 to correct PR field.)
CC (Updated on 27-AUG-2003 to correct OS field.)
XX SQ Sequence 708 AA;
Alignment Scores:
Pred. No.: 1,348-27 Length: 708
Score: 397.50 Matches: 181
Percent Similarity: 38.88% Conservative: 125
Best Local Similarity: 23.00% Mismatches: 270
Query Match: 10.65% Indels: 211
Gaps: 33
US-10-625-972-4 (1-2091) x AAY34501 (1-708)
Qy 40 CTGGATTTTCAGCCAGCAGCATAGCTGCTGAGAGGATGTG----- 81
Db 11 IleGlyIleSerAlaGlnAlaPheAlaLysThrAspAsnValProThrAspSerLeuArg 30
Qy 82 -----ATGATTGCTCGGATCGGCTATGAGAAAAGCTGACTAAC 123
Db 31 ValHisAsnLeuGlnThrValThrValTyrSerThrArgThrAlaValProLeuLysLys 50
Qy 124 GCAGCCCGCAGTGTCTTCTGTGATTAGCCAGAGGAAATTCAGTCCAGCCAGTACCAGAT 183
Db 51 IleProAlaLysMetGluLeuIleSerSerArgAsnIleLysGlnSerGlyPheAsnAsn 70
Qy 184 CTGGCGGAGGCTCTGAGATCAGTAGAGGCTGGATGTTGAAGTGGTACGGGTAAACC 243
Db 71 MetThrAspIleLeuLysThrGlnSerSerLeuAspValIleGlnTyrProGlyPheSer 90
Qy 244 GGAGGGCTGGAATCAGCATCCGAGGAATGCCAGCCAGT-----TACAGCTGATACTG 297
Db 91 Ser-----AsnIleGlyIleArgGlyPheLysProSerGlyLysTyrValThrValLeu 108
Qy 298 ATTATGTTGTTCTGTCAGCGCGGAAGCAGTACGCTGACTCCCAACGGTTTTTCTGCCATG 357
Db 109 ValAsnGlyIleProAlaGly-----ThrAspAsnIleSerThrLeu 122
Qy 358 AATACCGGTTTCATGCCCTCTGCGCCGATTCAGCTATTGAGGTATTCAGGGGCGCG 417
Db 123 AsnThr-----SerAsnIleGluGlnIleLeuLeuLysGlyPro 136
Qy 418 ATGTCCACACTGTATGCTCTGATGCGATGGCGGTGTGTGTATATCATATACAGAAAG 477
Db 137 PheSerSerIleTyrGlyThrAsnAlaMetGlyValValAsnIleIleThrHisLys 156
Qy 478 ATGCGACAGAAA-----TGCTC 495

Db 157 SerLysAspLysIleHisGlyAsnValSerLeuPheGlySerTyrGlnThrMetAla 176
 QY 496 TCTTCGTCATATCCAGG-----CTGAAT 519
 Db 177 GlySerPheAsnLeuGlyGlyArgPheGluAspIlePheSerPheAspLeuSerLeuGly 196
 QY 520 CTCAGAGAAAGCAACAAATGGGTACACAGCAGCCAGTTAAATTCCTGGAGCAGTGGTCCC 579
 Db 197 LeuAspLysGlnAsnLysAspTyrLysThrGlySerAsnAsnPheLeuSerLeuSerLys 216
 QY 580 CTGTGTGATGATCTCTAGCTGACGTGACGTGACGTGACGTGACGTGACGTGACGTGACGTGAC 639
 Db 217 LeuGluGluAlaIleValAspValAsnAlaThr---LysAsnLysLysMetLysGlySer 235
 QY 640 TCGGTCCATCATCTAGCAGCATACAGCAGCAGCGCTATTCCTATCCACAGGAG----- 693
 Db 236 AspTyrThrValAlaThrGlyArgLeuArgPheGlyIleAspPheThrProGluTrpSer 255
 QY 694 -----TCAGAAATTAATCTTGTGTGACGTCTTGTGACGTGACGTGACGTGACGTGACGTGAC 744
 Db 256 LeuAsnLeuTyrGlnAsnValPheLeuGlyAspAlaIleProValGly----- 271
 QY 745 GATGTCTCTGTTTCATATGATACCCAGCGGAGCGGTATGATACCGGATGGGCAA 804
 Db 271 ----- 271
 QY 805 CTGGGAGTCTGACGGGGGATATGACCGGACCTCGCTATGACGGAACAAATTTCA 864
 Db 272 ---GlySerIleTrpGlyValTyrGlyGluSer---LysLysAsnLeuAsnArgSerSer 289
 QY 865 GCTGGCTATGAT-----CATCTTTCCACCTTCGGACACATGGAAA 903
 Db 290 ThrSerPheGluLeuLeuGlyLysHisGlyCysHisThrLeuGlnPhe-----Ser 306
 QY 904 TCGTATCTCACTGGAACGACAGACAGAAAT----- 933
 Db 307 ProTyrPheAsnIleGluLysSerGluAsnTyrAsnAsnAlaAspProThrGlyPheIle 326
 QY 934 -----AAAGTCTGAGTCTTACGAGTGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 987
 Db 327 AsnTyrLysSerAspTyrTyrThrGlyAlaLeuLeuGlnAspLysIleSerPheGly 346
 QY 988 GTTCAGCCCGGAGCTTAAGGAATCGAACCTTATCTCTGAATTCATTACTGTTACCCCT 1047
 Db 347 GlyGln-----AsnIleValLeuGlyVal----- 354
 QY 1048 CTGGGAGATCTCATCTGTTACGTGGGGCGAG---TTTCAGAGCTCGCTCCATGAAA 1104
 Db 355 -----AspSerArgAsnMetThrMetGluSerGluArgPheGluGlnAla----- 369
 QY 1105 GACGAGTTGCTCTGCCAGCAGCAGGTGAACCTTCCCGGAGAAAAGCTGCTCGGTATTT 1164
 Db 370 ---GlyValAsnThrLysProTyrAsnProGlyTyrAlaThrAsnAsnIleGlyLeuPhe 388
 QY 1165 GCTGAGGATGATGGCAT---CTCAGGATGACCTTGGCTGACTGCGGGAGCGGC--- 1218
 Db 389 GlyGlnAlaAsnPheTyrLeuLeuAsnAspAlaLeuSerIleSerAlaGlyAlaArgAla 408
 QY 1219 -----TATGAACATCATGAGCAA 1236
 Db 409 AspPheMetPhePheAspLeuLysAlaAsnGluTyrLeuAsnAsnGluAlaLysGlnGlu 428
 QY 1237 TCCGGGGACACTTCAGTCCCGGTGCATATCTGCTGGGATGTCGAGATGCTCGGACG 1296
 Db 429 ThrHisAsnValIleAsnProAsnValGlyIleLysTyrGluPheValLysGlyLeuThr 448
 QY 1297 CTGAAAGCGGTGTGACCGGGATATAAGGACCCAGATGGGGCAGCTCATATAAGGG 1356
 Db 449 AlaHisGlyThrPheGlySerAlaPheSerAlaPheAspAlaPheGln-----LysAla 466
 QY 1357 ATTAGTGTGTCTCCGGCAGGAGAAACAAATCTTACTTGTGAACCCGACCTCGAAGCCG 1416
 Db 467 GlyGlnTyrValGlyProPheGlyThrThr-----IleGlyAsnProAspLeuLysPro 484

QY 1417 GAACAGAGCGTCAGTTATGAGGCTGGGTGTATTACGATAACCC---GCCGCTCGAAT 1473
 Db 485 GluLysSerMetThrTrpAspPheGlyIleGlyTyrSerAsnAlaArgCysGlyIleGln 504
 QY 1474 GCCAATGTC-----ACAGGTTTTATGACTGACTTCTCCAAACAAGATTGTCTCT--- 1521
 Db 505 AlaAspValThrLeuThrTyrPheHisThrAspHisLysAspLeuIleLeuSerSerPro 524
 QY 1522 ---TATTCATAAATGATACACCAATAGCTATGTAAACAGCGGAAAGCCCGTTGTCAC 1578
 Db 525 AspTyrAla-----AsnAsnIleThrThrTyrIleAsnAlaAspLysAlaArgMetSer 542
 QY 1579 GGTGTGAA-----TTTCCCGGCACATTGCCGCTG 1608
 Db 543 GlyIleGluAlaLeuLeuSerTyrAspPheGlySerLeuPheAlaAsnLysPheSerLeu 562
 QY 1609 TGTGCA---GAGGATGTCAGCTGTCTCACTCAATTAACCTGACCGCAAGTGAACAACGT 1665
 Db 563 ArgAlaPheAlaAsnAlaThrIleMetLeuAsnSerGluMetLysLysSerGlnThr--- 581
 QY 1666 GATGGTCAACAAAGGTGCGCGCTGAGTTATACCCCTGAACACATGCTGTAATGCGAAA 1725
 Db 582 -----AspAlaProTrpSerGluMetTyrTyrValArgLysGlnAsnIleThrPheGly 599
 QY 1726 CTGAACCTGGCAGATACCGAAGAGTGGCATCATGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 1785
 Db 600 IleGluTyrArgGlyLysGlyGlyLeuGluValMetLeuAsnGlyArgPheMetGlyArg 619
 QY 1786 ACACCACTTACCCAGAAATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1845
 Db 620 -----ArgIleGluGlnAsnTrp-----TyrAlaTyrTyrProGluValArgPro 634
 QY 1846 GAGAAAGGAGATACCTGAAAGCC----- 1869
 Db 635 GluLeuGlnLeuLeuAlaGluGluProGluLeuAlaGlnGlyLeuLeuArg 654
 QY 1870 -----TGGACGGTGGTGGATGCGGTCTGCTGGAAGATGACGATCCCTGACG 1920
 Db 655 HisProGlnAlaMetValPheAsnAlaSerAlaTyrHisMetAsnLysTyrLeuThr 674
 QY 1921 CTGAATGCTGCGGTGAATTAACCTGCTCAACAAGATTACAGTACGCTGACGCTGCTGACGT 1980
 Db 675 PheGlyValAsnLeuAsnAsnIleLeuAspGluLeuTyrThrGlu----- 689
 QY 1981 GCGGTAAGAGTACGCTGATGCGGTGATTACTTCCAGACGGGATCATCAACAACAGGA 2040
 Db 690 -----LysAspGly 692
 QY 2041 TATGTGATACCTGACGGAAT 2061
 Db 693 TyrHisMetProGlyArgAsn 699
 RESULT 11
 ID AAY34374
 XX AAY34374 standard; protein; 772 AA.
 AC AAY34374;
 DT 27-AUG-2003 (revised)
 DT 20-MAR-2003 (revised)
 DT 25-AUG-1999 (first entry)
 DE Porphyromonas gingivalis protein PG40.
 XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis; vaccine;
 KW antigenic.
 OS Porphyromonas gingivalis.
 XX W09929870-AL.
 XX 17-JUN-1999.


```

QY 1237 TTCCGGGGACACTTCAGTCGGCTGCATATCTGCTGGGATGTCGCAGATCCCTGGACG 1296
Db 493 ThrHisAsnValIleAsnProAsnValGlyIleLysTyrGlupheValIysGlyLeuThr 512
QY 1297 CTGAAAGGGGGTGTGACCCAGGATATAAGGCACCCAGAAATGGGAGCATACATAAAGG 1356
Db 513 AlaHisGlyThrPheGlySerAlaPheSerAlaProAspAlaPheGln-----LysAla 530
QY 1357 ATTAGTGGTGTCCGGGAGGAGGAAAAACAATCTACTTGGTAACCCGACCTGAAGCG 1416
Db 531 GlyGlnTyrValGlyProPheGlyThrThr-----IleGlyAsnProAspLeuLysPro 548
QY 1417 GAAGACAGGCTCAGTTATGAGTGGGTGTATTACGATAACCC-----GCCGGTCTGAAT 1473
Db 549 GluLysSerMetThrThrPhePheGlyIleGlyTyrSerAsnAlaArgCysGlyIleGln 568
QY 1474 GCAATGTC-----ACAGTTTATGACTGCTCTCCAAACAAGATTGCTCT----- 1521
Db 569 AlaAspValThrLeuThrTyrPheHisThrAspHisLysAspLeuIleLeuSerSerPro 588
QY 1522 ---TATTCATAAATGATACACCAATAGTATGTAACAGCGGAAAGCCCGGTGCAC 1578
Db 589 AspTyrAla-----AsnAsnIleThrThrTyrIleAsnAlaAspLysAlaArgMetSer 606
QY 1579 GGTGTGGAA-----TTTCCCGGCACATGTCGGCTG 1608
Db 607 GlyIleGluAlaLeuLeuSerTyrAspPheGlySerLeuPheAlaAsnLysPheSerLeu 626
QY 1609 TGGTCA--GAGATGTCAGCTGTACCTGAATTACCTGGACCCGAGTGAACAACGT 1665
Db 627 ArgAlaPheAlaAsnAlaThrIleMetLeuAsnSerGluMetLysLysSerGlnThr--- 645
QY 1666 GATGGTGATAACAAAGTGGCGGTGAGTTATACCTGTAACACATGCTGAATCGGAA 1725
Db 646 -----AspAlaProTrpSerGluMetTyrTyrValArgLysGlnAsnIleThrPheGly 663
QY 1726 CTGAACCTGGCAGATCACCGAAGGTGGCATCATGCTGGTCCCGTTCGCGGAAA 1785
Db 664 IleGluTyrArgGlyLysGlyLeuGluValMetLeuAsnGlyArgPheMetGlyArg 683
QY 1786 ACACCAAGTTTACCCAGAAATTATTCGTCACTGAGCGGTGTACAGAAAGATGTATGAT 1845
Db 684 -----ArgIleGluGlnAsnTrp-----TyrAlaTyrTyrProGluValArgPro 698
QY 1846 GAGAAAGGAGATPACTCGAAAGC----- 1869
Db 699 GluLeuGlnGlnLeuLeuAlaGluProGluLeuAlaAlaGlnGlyLeuLeuArg 718
QY 1870 -----TGGACGTGTGGTGGATCGAGTCTGTCTGGAAGATGACGGATCCCTGACG 1920
Db 719 HisProGlnAlaValMetValPheAsnAlaSerAlaTyrTyrHisMetAsnLysTyrLeuThr 738
QY 1921 CTGAATGCTCGGTGAATACCTGCTCAACAAGGATTACAGTACGCTGAGCTGTACAGT 1980
Db 739 PheGlyValAsnLeuAsnIleLeuAspGluLeuTyrThrGlu----- 753
QY 1981 GCGGTAAAGATACGTGTATCCGGTGTATCTCCAGACGGGATCATCAACACAGGA 2040
Db 754 -----LysAspGly 756
QY 2041 TATGTGATACCTGAGCGAAAT 2061
Db 757 TyrHisMetProGlyArgAsn 763

```

RESULT 12

ADCC01296

ID ADC01296 standard; protein; 660 AA.

XX AC ADC01296;

XX AC ADC01296;

DT 04-DEC-2003 (first entry)

DE Enterohaemorrhagic *E. coli* O157:H7-specific protein SEQ ID NO: 1340.

XX enterohaemorrhagic; anti-bacterial.
 KW Escherichia coli; O157:H7.
 OS JP2002355074-A.
 PN 10-DEC-2002.
 PD 24-JAN-2002; 2002JP-00015959.
 PF 24-JAN-2001; 2001JP-00112010.
 PR (UITS-) UNIV TSUKUBA.
 PA WPI; 2003-451640/43.
 DR Enterohaemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule and a polypeptide and its use, a polypeptide, a vector and a host cell.
 PS Claim 3; SEQ ID NO 1340; 2067pp; Japanese.
 CC The invention relates to a novel enterohaemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule. A polynucleotide of the invention has anti-bacterial activity. The polypeptide can be used in detection and/or treatment of O157:H7 infection. The nucleotide sequence of the genome of Enterohaemorrhagic *E. coli* O157:H7 was determined. The present sequence represents an *E. coli* O157:H7-specific polypeptide of the invention.
 CC Sequence 660 AA;
 SQ

Alignment Scores:

Pred. No.: 5,21e-27 Length: 660
 Score: 391.00 Matches: 183
 Percent Similarity: 39.16% Conservative: 106
 Best Local Similarity: 24.80% Mismatches: 257
 Query Match: 10.48% Indels: 192
 DB: 36 Gaps: 36

US-10-625-972-4 (1-2091) x ADC01296 (1-660)

QY 1 ATCGGATAACCACTCTGGCT-----TCGTAGTCATTCCTCTGTCGGATTTC 51
 Db 9 LeuArgLeuSerLeuLeuAlaLeuAlaValSerAlaThrLeuProThrPheAlaPhe--- 27
 QY 52 GCCAGCAGCATAGCTGCTGCAGAGGATGTCATGATTGTCGCGCATCCGCTATGAGAA 111
 Db 28 -----AlaThrGluThrMetThrValThrAlaThrGlyAsnAlaArg 41
 QY 112 AAGCTGACTAACCCAGCGCCAGTGTCTGTGATTAGCCAGGAGGAATGCGAGTCCAG 171
 Db 42 SerSerPheGluAlaProMetMetValSerValIleAspThrSerAlaProGluAsnGln 61
 QY 172 CAGTACCACGATCTGCGGAGGCTCTGAGATCAGTAGAGGCTGTGATGTTGAAAAGTGT 231
 Db 62 ThrAlaThrSerAlaThrAspLeuArgHisValProGlyIleThrLeuAsp---Gly 80
 QY 232 ACGGTAACACCGAGGCTGGAATCAGCATCCGAGGAATCCAGCCAGTTACAGCTG 291
 Db 81 ThrGlyArgThrAsnGlyGlnAspValAsnMetArgGlyTyrAspHisArgGlyValLeu 100
 QY 292 ATACTGATTGATGCTGTCTGTCAGGCGGAGGAGCAGTGACGTCACTCCCAACGGTTTTCT 351
 Db 101 ValLeuValAspGlyValArgGlnGly-----ThrAspThrGlyHis--- 114
 QY 352 GCCATGAATACCGGTTTCATGCCCTCTGCGCGCCATGAGCGTATGAGTTATCAGG 411
 Db 115 ---LeuAsnGlyThrPheLeuAspPro---AlaLeuIleLysArgValGluIleValArg 132
 QY 412 GGGCCGATGTCACACTGTATGCTCTGATGGATGGGCGGTGTGGTGAATCATCATACC 471
 Db 133 GlyProSerAlaLeuLeuTyrGlySerGlyAlaLeuGlyValIleSerTyrAspThr 152

XX WPI; 2001-550253/61.
 XX A library of DNA fragments of *Escherichia coli* strains for the phylogenetic
 PT determination of a given strain comprises polynucleotides of nature B2/D+
 PT A-.
 XX Example 6; Fig 6; 646pp; English.
 XX The invention relates to a library of DNA fragments of *Escherichia coli*
 CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533) and
 CC encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
 CC B2/D+A-. The polynucleotides have potential anti-inflammatory,
 CC antibacterial and immunosuppressive activity as part of pharmaceutical
 CC compositions used to treat, palliate or prevent extra-intestinal *E. coli*
 CC infections. The polypeptides are useful for determining the phylogenetic
 CC group of a given *E. coli* strain. These polypeptides can detect and treat
 CC an undesired development of *E. coli*, particularly an extra-intestinal
 CC infection that include systemic and non-diarrhoeal infections such as
 CC septicemia, pyelonephritis and meningitis this is particularly
 CC advantageous as bacterial resistance is increasing with the more frequent
 CC use of broad spectrum antibiotics
 XX SQ Sequence 660 AA;

Alignment Scores:
 Pred. No.: 9,94e-27 Length: 660
 Score: 388.00 Matches: 183
 Percent Similarity: 40.17% Conservative: 107
 Best Local Similarity: 25.35% Mismatches: 272
 Query Match: 10.40% Indels: 160
 DB: 4 Gaps: 37

US-10-625-972-4 (1-2091) x ABB52669 (1-660)

XX	DB	162	-----GlyGlnSerSerGlyPheArgValPheGlyThrGlyGlyThrGlyAspHis	178
DR	QY	592	TCTCTACGCTG-----CAGGTACGCGTAGCACCAACAGCGTCCAGGTTTCATCGGTC	645
XX	DB	179	SerLeuGlyLeuGlyAlaSerAlaPheGlyArgThrGluAsnLeuAspGlyIleValAla	198
PT	QY	546	ACATCACTGAGCATACAGCAGGCGCCGT-----ATTCTTATCCACGGAG	693
XX	DB	199	TrpSerSerArgAspArgGlyAspLeuArgGlnSerAsnGlyGluThrAlaProAsnAsp	218
CC	QY	694	TCACAGAAATATATCTTGTGTCACGCTCTTGAAGGCGGTCGAGCAGGAT-----	747
CC	DB	219	GluSerIleAsnAsnMetLeuAlaLysGlyThrTrpGlnIleAspSerAlaGlnSerLeu	238
CC	QY	748	-----GTGCTCTGCTTGTATATGATACC-----ACCCGG	777
CC	DB	239	SerGlyLeuValArgTyrTyrAsnAsnAspAlaArgGluProLysAsnProGlnThrVal	258
CC	QY	778	CAGCGTTATCATAAACCGGATGGCAACTGGGAGTCTGACGGGGGATATGACGGACC	837
CC	DB	259	GluAlaSerAspSerSerAsnProMetValAspArgSerThrIleGlnArgAlaGln	278
CC	QY	938	CTGCGTATGAGCGAAACAAAATTCAGTGGCTATGATCATCTTTCACCTTCGGAACA	897
CC	DB	279	LeuSerTyrLys-----LeuAlaProGlnGlyAsnAsp-----	289
QY	QY	898	TGG-----AAATCGTATCTGAACGAGCAGACAGAAAATAAGTCTGAGCTT	948
DB	DB	290	TrpLeuAsnAlaAspAlaLysIleTyrTrpSerGluValArg-----	303
QY	QY	949	GTACGCGTGTACTGAAGCGCGCAAAATGGCGCTTCGCGTCAGCGCGGAG-----	1002
DB	DB	304	-----IleAsnAlaGlnAsnThrGlySerSerGlyGluTyrArgGluGlnIle	319
QY	QY	1003	CTTAAGAAATCGAACCTTATCTGAATTCATCTACTCTTACCCCTCGGAGAACTCAT	1062
DB	DB	320	ThrLeuGlyAlaArgLeuGluAsnArgSerThrLeuPheAlaAspSerPheAlaSerHis	339
QY	QY	1063	CTGTTACGCTGGCGGCGAGTTTCAGAGCTCGTCCATGAAGACGCGAGTTCTCTTGC	1122
DB	DB	340	LeuLeuThrTyrGlyGlyGluTyrTyrArgGlnGluGlnHisProGlyGly-----	357
QY	QY	1123	AGCAGAGTGAACCTTTCGCGCAGAAAGC-----TGTCTGATTTCTGCTGAGAT	1173
DB	DB	358	ThrThrGly-----PheProGlnAlaLysIleAspPheSerSerGlyTrpLeuGlnAsp	375
QY	QY	1174	GAGTGCATCTCACGGAT---GCACCTTGCCTGACTCGCGGCGAGCGCTATGAACATCAT	1230
DB	DB	376	GluIleThrLeuArgAspLeuProIleThrLeuLeuGlyThrArgTyrAspSerTyr	395
QY	QY	1231	GAGCAATTCGGGGACACTTCAGTCGCGTGCATATCTCGTCTGGGATGTG---GCAGAT	1287
DB	DB	396	ArgGlySerSerAspGlyTyrLys-----AspValAspAlaAsp	408
QY	QY	1288	GCCTGACGCTGAAGCGCGTGTGACC-----	1314
DB	DB	409	LysTrpSerSerArgAlaGlyMetThrIleAsnProThrAsnTrpLeuMetLeuPheGly	428
QY	QY	1315	-----ACGGATATAAGCACCAGAAATGGGCGAGCTACATAAAGGGATTAGTGT	1365
DB	DB	429	SerTyrAlaGlnAlaPheArgAlaProThrMetGlyGluMetTyrAsnAspSerLysHis	448
QY	QY	1366	GTGTCC---GGCAGGGAACAAACAACTA---CTTGGTAACCGCAGCTGAGCGGAA	1419
DB	DB	449	PheSerIleGlyArgPheTyrThrAsnTyrTrpValProAsnProAsnLeuArgProGlu	468
QY	QY	1420	-----GAGAGCTCAGTTATGAGCTGGGTGTATTACGAT-----AAC	1458
DB	DB	469	ThrAsnGluThrGlnGluTyrGlyPheGlyLeuArgPheAspAspLeuMetLeuSerAsn	488
QY	QY	1459	CCGCGCGTCTGAATGCCAATGTACAGGTTTATGACT-----GACTTCTCCAAC	1509
DB	DB	489	AspAla---LeuGluPheLysAlaSerTyrPheAspThrLysAlaLysAspTyrIleSer	507


```

QY 1510 AAGATTGCTTATTCCATAAATGATAACACCAATAGTATGTAAACAGCGAAAGGCC 1569
Db 508 ThrThrValAspPheAlaAlaAlaThrThrMetSerTyrAsnValProAsnAlaLys--- 526
QY 1570 CGGTTGACGGTGTGGAATTGCGGCACATTCGCCGTGTGGTCAGAGGATGTC----- 1623
Db 527 ---:---:---:---:---:---:---:---:---:---:---:---:---:---:---:--- 534
QY 1624 -----ACGCTGTCACTGAATTACACCTGGACCGACCGAGTGAACACAGTGAT 1668
Db 535 LysTyrThrThrAspLeuPheSerLeuAspValAlaTyrAsnArgThrArgGlyLysAsp 554
QY 1669 GGTGATAACAAAGTGGCGCGCTGAGTTATACCCCTGAACACATGCGTAATGCGAAACTG 1728
Db 555 ThrAspThr---GlyGluTyrIleSerSerIleAsnProAspThrValThrSerThrLeu 573
QY 1729 AACTGGCAGATCACCGAGAGGTGGCATCA-----TGGCTGGGTGCCCTTATCGCGGG 1782
Db 574 AsnIleProIleAlaHisSerGlyPheSerValGlyTyrValGlyThr---PheAlaAsp 592
QY 1783 AAAACACACAGCTTTACACCCAGAAATTATTTCGTCACTGAGCGCTGTACAGAGAAAGTGAT 1842
Db 593 ArgSerThrHisIleSerSerTyrSerLys----- 603
QY 1843 GATGAGAAAGAGAAATACCTGAAGCCCTGGACGGTGTGGATGACAGCTCTGTCTGGAAG 1902
Db 604 -----GlnProGlyTyrGlyValAsnAspPheTyrValSerTyrGln 617
QY 1903 ATGACGGATGCCCTG-----ACGCTGAATGCTGGGTGAATAACCTGCTCAACAAG 1953
Db 618 GlyGlnGlnAlaLeuLysGlyMetThrThrThrLeuValLeuGlyAsnAlaPheAspLys 637
QY 1954 GATTAC 1959
Db 638 GluTyr 639

RESULT 14
ADA33859
ID ADA33859 standard; protein; 643 AA.
XX AC
XX ADA33859;
XX
DT 20-NOV-2003 (first entry)
DE Acinetobacter baumannii protein #1020.
XX
XX Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
KW plant biocontrol agent.
XX
OS Acinetobacter baumannii.
XX
XX US6562958-B1.
XX
XX 13-MAY-2003.
XX
XX 04-JUN-1999; 99US-00328352.
XX
XX 09-JUN-1998; 98US-0088701P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Breton G, Bush D;
XX
XX WPI; 2003-576092/54.
XX
XX N-PSDB; ADA29733.
XX
XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
XX for diagnosing a bacterial disease, as components of antibacterial
XX vaccines, as targets for antibacterial drugs, or as biocontrol agents for
XX plants.
XX
XX Example; SEQ ID NO 5146; 328pp; English.
PS

```

```

XX The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents the amino acid sequence of an A.
CC baumannii protein.
XX
XX Sequence 643 AA;
SQ
Alignment Scores:
Pred. No.: 2,598-26 Length: 643
Score: 383.50 Matches: 169
Percent Similarity: 38.61% Conservative: 109
Best Local Similarity: 23.47% Mismatches: 279
Query Match: 10.28% Indels: 163
DB: 6 Gaps: 32
US-10-625-972-4 (1-2091) x ADA33859 (1-643)
QY 10 ACCACTCTGGCTTCGGTAGTCAATCCCTGCTCTCGGATTTTCAGCAGCAGCATACTGCT 69
Db 31 ThrArgLeuValGlyAlaIleAlaIleAlaMetGly-----CysSerProValIlePhe 48
QY 70 GCAGAGGATGTG-----ATCATTGCTCGCATCCCGCTAT 105
Db 49 AlaGluAspAlaThrAspAlaThrGlnLeuAspProIleValIleThrAlaSerLysSer 58
QY 106 GAGAAACCTGACTAACCGCAGCGCCAGTGTCTGTGATTAGCCAGGAGAAATTCAG 165
Db 69 AlaGluLysAlaSerGluValProAlaArgIleSerIleGluProLysIleValGlu 88
QY 166 TCAGCCAGTACCATCTCGCGAGGCTCTGAGATCAGTAGAGGCTGTGATGTT--- 222
Db 89 GlnSerProIleAlaGluLeuProHisLeuLeuMetSerAspAlaAlaIleAsnMetVal 108
QY 223 ---GAAAGTGTACGGGTAAACCGGAGGCTGGAATCAGATCCGAGGAATGCCAGCC 279
Db 109 GlnSerGlyGlyLeuGlyGlnThrSer-----SerIlePheIleArgGlyThrAsnSer 126
QY 280 AGTTACACCGCTGATCTACTGATTGATGTTCTGTCAGGCGGAGAGAGTACGCTGACTCCC 339
Db 127 GluHisAlaLeuIleLeuArgAspGlyAlaArgLeuAsnThrAlaSerThr----- 143
QY 340 AACGGTTTTCCTCATGATACCGGGTTCATGCCCCCTCTGGCCGCCATTCAGCGTATT 399
Db 144 -----GlyAlaAlaAsnLeuAlaPheIle---AspThrThrAspLysIleGlnIle 159
QY 400 GAGGTTTATCAGGGGCGCGATGTCCACACTGTATGGCTCTGATCGCATGGCGGTGGTG 459
Db 160 GluIleLeuLysGlyProAlaSerValLeuTyrGlyThrAspAlaIleGlyGlyValVal 179
QY 460 AATATCATTTACAGAAAGATGACAGAAATGGCTCTCTTCGTCATTCAGCGGCTGAAT 519
Db 180 GlnIleIleSer---LysThrProGluLysThrSerAlaPheValThrGlyGluIle--- 197
QY 520 CTGCAGGAACCAACAAATGGGTAAACAGCAGCAGTCTTAATTTCTGGAGCAGTGTGCC 579
Db 198 -----GlyGluAsnLysThrTyrLysSerIleValGlyAlaAsp 210
QY 580 CTTGTGGATGATTCTGTCAGCCTGACGGTACGGGTAGCACACACAGCGTTCAGGGTTCA 639
Db 211 LeuAlaGluAsnGlyPheTyrAlaGlnValArgGlyGlnArgLeuGluSerAspGlySer 230
QY 640 TCGGTACATCACTGAGC-----GATACAGCAGGACG 672
Db 231 ArgIleThrAspLeuLysGlyAsnAspIleLysLysAlaSerTyrAspGlnLysGlyPhe 250
QY 673 CGTATTCCTTATCCCGAGTACAGAAATATATCTTGTGTCACGCTCTTGACTGGAAG 732

```

The invention relates to isolated Acinetobacter baumannii nucleic acids. The A. baumannii nucleic acids and polypeptides are useful as reagents for diagnosing a bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, to detect the presence of A. baumannii and other Acinetobacter species in a sample, in screening compounds for the ability to interfere with the A. baumannii life cycle or to inhibit A. baumannii infection, and as biocontrol agents for plants. The present sequence represents the amino acid sequence of an A. baumannii protein.

CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi that
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which P.
 CC luminescens is a model (particularly plague and whooping cough). This
 CC sequence represents one of the isolated P. luminescens proteins
 XX
 SQ Sequence 674 AA;

Alignment Scores:
 Pred. No.: 2,27e-25 Length: 674
 Score: 373.50 Matches: 174
 Percent Similarity: 39.56% Conservative: 131
 Best Local Similarity: 22.57% Mismatches: 273
 Query Match: 10.01% Indels: 193
 DB: 6 Gaps: 36

US-10-625-972-4 (1-2091) x ABM67198 (1-674)

QY 7 ATACCACTCTGGTTCGCTAGTATCCCTGCTCGGATTTTCAGCCAGCAGCATAGCT 66
 Db 19 lIeSerAsnIleProAlaValTyAlaAlaThrLysSerProAspAlaAsnSerAspLys 38
 QY 67 GCTCAGAGGATGATGATGCTCGGCATCCCGCTATGAGAAAGCTACTAACGCA 126
 Db 39 SerSerArgGluValIleThrValTyAlaThrGlyAsnGluArgSerPheGluAla 58
 QY 127 GCGCCAGTGTTCGTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCAAGTCTG 186
 Db 59 ProMetMetValThrValIleAsnSerGlnSerProGlnSerGlnThrAlaGlyAsnAla 78
 QY 187 GCGAGGCTCTGAGTACGATGAGGCTGTGATGTTGAAAGTGTGACGGGTAAACCGGA 246
 Db 79 AsnAspLeuIleLysIleProGlyIleAsnIleAlaGlyIleAsnArgAlaAsn 97
 QY 247 GGGCTGGAATCAGCATCCGAGGATGCCAGCAGTTACCGCTACGATGATGATGGT 306
 Db 98 GlyGlnAspValSerLeuArgGlyTyGlyProLysGlyValLeuThrLeuIleAspGly 117
 QY 307 GTTCGTCAAGGCGGAAGCAGTGCATGCTCCCAACGGTTTTTCTGCGCATGAATACCGGG 366
 Db 118 lIeArgGlnGly-----ThrAspThrGlyHis-----IleAsnGlyThr 130
 QY 367 TTCATGCCCCCTCTGCGCCCATTCAGCGTATTGAGGTATACAGGGCCGATGCCACA 426
 Db 131 PheLeuAspPro---AlaLeuIleLysGlnValGluValAlaArgGlyProSerAlaAla 149
 QY 427 CTGTATGCTCTGATGCGTGGCGGTGTGTAATATCATTACCAAGAAAGATCCAGAC 486
 Db 150 LeuTyGlySerGlyAlaLeuGlyGlyValIleAlaTyGlnThrValAspAlaAlaAsp 169
 QY 487 AAATGGCTCTCTCCGTCATGCAAGGCTGAATCTGCAGGAAGCAACAATGGGTAAAC 546
 Db 170 LeuLeuGlnAlaGlyGlnAspHisGlyPheArgValPheAlaArgGlyAlaThrGlyAsp 189
 QY 547 AGCAGCCAGTTAAATTC-----TGAGCAGTGTGCTCCCTGTGCGATGATCT 594
 Db 190 HisSer---MetGlyPheGlyAlaThrPheAlaLysSerGluSerLeuAspGlyVal 208
 QY 595 GTCAAGCTCAGGTACGCGGTAGCACACACAGCGTTCAGGGTTCATCGGTACATCACTG 654
 Db 209 PheAlaPheSerThrArgAspSerGlyAsnIleArgHisGlyGlyValValGlyAsp 228
 QY 655 AGCGATACAGCAGCAGCGTATCTTATCCAGCGAGTACAGAAATTAATCTGGT 714
 Db 229 AsnAspGluPheIleSer-----GAGGAGCAGGATGCTC-----AsnLeuMet 237
 QY 715 GCACGCTCTGACTGGAAGCGTCGGAGCAGGATGCTC-----ACCGAAGTGAACAA 1662
 Db 238 AlaLysGlySerTrpLysIleAspAsnGlnSerGlnLeuSerSerGlnLeuArgTyTy 257
 QY 753 ----- 753

Db 258 ArgAsnGluAlaAsnGlnProLysAsnProGlnThrLeuThrGlyLysAsnLysPro 277
 QY 754 -----TGTTTCATATGATACCATCCCGCAGCGGTATGATAACCGGATGGCAACTG 807
 Db 278 SerThrTrpThrAspArgThrThrGln-----ArgAspAlaGlnLeu 292
 QY 808 GGGAGTCTGACGGGGGATATGACCGACCTCGCTATGAGCAACAAAATTTTCAGT 867
 Db 293 Ser-----TyGlnLeuAsn-----ProAla 299
 QY 868 GGCTATGATCATATCTTTCACCTTCGGAACATGG-----AAATCGTATCTGACTG 918
 Db 300 GluTyAsp-----TrpLeuAsnAlaLysThrAspLeuTyTy 312
 QY 919 AACGACACAGAAAATAAGGTCTGAGCTTGACGAGTGTACTGAAGCGCAGCAATGG 978
 Db 313 SerAspIleThrIleAsnAlaArgThrLysAlaLys----- 324
 QY 979 GGCTTGGCGCTCAGCGGGAGCTTAAGGAATCGAACCTTATCTGATTCATTACTG 1038
 Db 325 GlyPheGluGlyArgLysGlnLysThrTyGlyValLysLeuGluAsnArgSerArgLeu 344
 QY 1039 CTTACCCCTCTGGGAGAATCTCATCTGTTACGTTGGGGGGGAG-----TTTCAGAGCTCG 1095
 Db 345 TrpThrAspSerProLeuAlaHisGlnPheThrTyGlyGlyGluThrTyLysGlnLys 364
 QY 1096 TCCATGAAGAAGCAGAGTGTCTTCCAGCAGCAGGTGAACCTTCCGGCAG----- 1146
 Db 365 GlnThrProAspGly-----AsnThrAspSerPheProAspAlaAspIle 379
 QY 1147 AAAAGCTGGTGGTATTTCGTGAGGATGAGTGCATCTCAGGAT---GCACCTGGCGTG 1203
 Db 380 ArgPheTySerGlyTrpLeuGlnAspGluIleThrLeuArgAspLeuProValSerIle 399
 QY 1204 ACTCGCGGACCGCTATGAACATCATGACAACTCGGGGACACTTCAGTCCGCTGCA 1263
 Db 400 ValAlaGlyThrArgTyAspAsnTyLys-----GlyAsnAsnSerLysAspGluAla 417
 QY 1264 TATCTGCTGGGATGGCAGATGCTGAGCCTGAAAGCGGTGTGACC----- 1314
 Db 418 Val-----SerAlaAspLysTrpSerSerLysGlyAlaIleSerIleThrPro 433
 QY 1315 -----ACGGGATATAAGCACCACCGAATGGGG 1341
 Db 434 ThrAspTrpSerMetLeuPheThrSerTyAlaGlnAlaPheArgAlaProThrMetGly 453
 QY 1342 CAGCTACATAAAGGATAGTGTGTCTCCGCGCAGGAAACAACTACTTCTGGTAA 1401
 Db 454 GluMetTyAsnAspSerMetHisTyProGlyAsnPheTrpThr-----ProAsn 470
 QY 1402 CCGAGCTGAAGCGGAGAGCGTCAGTTATGAGCTGGG----- 1443
 Db 471 ProAsnLeuArgProGluSerAsnGluThrTrpLysSerGlyPheGlyLeuArgPheAsn 490
 QY 1444 ---GTGTATTACGATAACCCCGCTGTAATGCCAATGTACAGGTTTATGACTGAC 1500
 Db 491 AspLeuLeuAlaAspAsnAspGlu---LeuLysPheLysAlaSerTyPheAspThrLys 509
 QY 1501 TTCTCCACAGATGTTCTCTTATTCATTAATGATAACACCAAT---AGCTATGTAAC 1557
 Db 510 AlaLysAspTyTrileLysSerAspIleValAsnArgGlyArgAsnThrTrpSerValAsn 529
 QY 1558 AGCGAAAGCGCGGTGTCAGCGTGTGAATTTTCGCGGACATTCGCGTGTGGTCAGAG 1617
 Db 530 lIeSerSerAlaLysIleTrpGly-----TrpAspAla 540
 QY 1618 GATGTCAGCTGTCATGTAAT---TACACCTGG-----ACCGAAGTGAACAA 1662
 Db 541 SerIleAsnTyGluSerAsnLeuPheSerTrpGluLeuAlaTyAsnArgThrGluGly 560
 QY 1663 CGTGATGCTGATACAAAGGTGGCGCGCTGAGTTATACCTCCATCAACACATGGTGAAT--- 1719


```
Db      561 LysAspGluLysSerGlyGlySerLeuAspAsnLeuSerProAspThrIleThrSerIle 580
QY      1720 -----GCGAAACTGAACTGGCAGATCACCGAAGAGGTGGCATCATGGCTG 1764
Db      581 LeuGluIleProIleAlaGlnThrGlyPheSerVal-----GlyTrpVal 595
QY      1765 GGTGCCCGGTATCGCGGAAACACACCGTTTCCACCCAGAAATTATTCGTCACTGAGCGCT 1824
Db      596 Gly-----GlnPheThr---AsnHisThrAspPheLysGly 606
QY      1825 GTACAGAAGAAAGTGTATGATGAGAAAGGAGAAATACCTGAAAGCCTGGACGGTGGTGGAT 1884
Db      607 LysAspAlaArgGlyArgAspProLysGlnGln-----AlaGlyTyrGlyValAsnAsp 624
QY      1885 GCAGGTCTGTCGGGAAG-----ATGACGGATGCCCTGACCGTGAATGCTCGGTG 1935
Db      625 PheTyrValSerTyrGlnGlyGluGlyMetLeuLysGlyValThrThrAlaValLeu 644
QY      1936 AATAACCTGCTCAACAAGGATTACAGTGACGTGACGTGAGCCTGTACAGTCCCGTAAAGAGTACG 1995
Db      645 GlyAsnAlaPheAspLysGluTyr----- 652
QY      1996 CTGTATGCCGGTGATTACTTCACAGCGGATCATCAACAGGATATGTGATACCTGAG 2055
Db      653 -----TyrSerProGlnGlyThrProGlnAspGly----- 662
QY      2056 CGAAATTACTGGATGTCGCTGAACTATCAGTTC 2088
Db      663 ArgAsnAlaLysLeuPheValSerTyrGlnTrp 673
```

Search completed: October 13, 2004, 11:48:58
Job time : 169 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2004, 11:56:38 ; Search time 5408 Seconds

(without alignments)
11546.194 Million cell updates/sec

Title: US-10-625-972-4

Perfect score: 2091

Sequence: 1 atgcgaataaacactctggc.....cgctgaactatcagttctga 2091

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba.*

2: em_esthm.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_hic.*

9: gb_est1.*

10: gb_est2.*

11: gb_hic.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rtd.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gsl.*

29: gb_gsl2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	286.8	13.7	290	9 AI613886	AI613886 vh77h07.y
2	278	13.3	289	9 AI481927	AI481927 vh86a12.x
3	251	12.0	495	28 A2935601	A2935601 T034 E.c
4	224.6	10.7	243	9 AA501138	AA501138 vh77h07.r

ALIGNMENTS

RESULT 1
AI613886

LOCUS

DEFINITION

vh77h07.y1 Knowles Solter mouse inner cell mass Mus musculus CDNA

clone IMAGE:893053 5', similar to TR:087518 O87518 EXOGENOUS FERRIC

SIDEROPHORE RECEPTOR R4. ; , mRNA sequence.

ACCESSION

AI613886

VERSION

AI613886.1

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

REFERENCE

AUTHORS

Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,

Underwood,K., Streptoe,M., Theising,B., Allen,M., Bowers,Y.,

Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,

Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,

Waterston,R. and Wilson,R.

TITLE

The WashU-NCI Mouse EST Project 1999

JOURNAL

Unpublished (1999)

COMMENT

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

AI613886 290 bp mRNA linear EST 21-APR-1999

vh77h07.y1 Knowles Solter mouse inner cell mass Mus musculus CDNA

clone IMAGE:893053 5', similar to TR:087518 O87518 EXOGENOUS FERRIC

SIDEROPHORE RECEPTOR R4. ; , mRNA sequence.

AI613886

AI613886.1 GI:4623053

EST.

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 290)

Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,

Underwood,K., Streptoe,M., Theising,B., Allen,M., Bowers,Y.,

Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,

Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,

Waterston,R. and Wilson,R.

TITLE

The WashU-NCI Mouse EST Project 1999

JOURNAL

Unpublished (1999)

COMMENT

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

5	81.8	3.9	596	28	AQ990279	Rfc01023
6	67.4	3.2	690	14	CF199793	EST1397 T
7	64.8	3.1	988	28	BZ576672	msh2-5047
8	62.4	3.0	1350	29	CG745412	P038-2-C0
9	57.2	2.7	519	28	AZ302660	GSSBtrul59
10	55.2	2.6	1029	28	BZ568287	pac82-164
11	54.8	2.6	351	29	CG591863	CG91863 0180518-0
12	51.6	2.5	598	14	CB287407	CB287407 CMD58-D06
13	51.6	2.5	602	14	CB287362	CB286719 CMD44-G05
14	51.6	2.5	609	14	CB286719	CB286521 CMD39-B07
15	51.6	2.5	618	14	CB286521	CB286442 CMD37-D12
16	51.6	2.5	637	14	CB286442	CB286061 BACPB21-O
17	51.6	2.5	683	28	BH825061	BH825061 BACPB21-O
18	48.2	2.3	1023	28	BZ557352	BZ557352 pac81-60
19	47.4	2.3	1082	28	BZ559361	BZ559361 pac82-164
20	46.8	2.2	1069	28	BZ577496	BZ577496 msh2-5436
21	46.8	2.2	1169	28	BZ565939	BZ565939 pac82-164
22	46.4	2.2	766	28	BH388728	BH388728 AG-ND-104
23	44	2.1	575	28	AQ989900	AQ989900 Rfc00567
24	43.8	2.1	1201	13	BX381961	BX381961 BX381961
25	42.8	2.0	322	10	AW356775	AW356775 39353 MAR
26	42.8	2.0	1188	28	BZ567160	BZ567160 pac82-164
27	42.6	2.0	872	28	BZ556928	BZ556928 pac81-60
28	42.6	2.0	1312	28	BZ575894	BZ575894 msh2-4675
29	42.4	2.0	1201	13	BX381961	BX381961 BX381961
30	41.8	2.0	575	14	CD087064	CD087064 MCL-0034T
31	41.4	2.0	643	13	BX381961	BX381961 BX381961
32	41.4	2.0	985	28	AF075901	AF075901 AF075901
33	41	2.0	924	28	BZ562000	BZ562000 pac82-164
34	40.8	2.0	1211	28	BZ579322	BZ579322 msh2-6302
35	40.4	1.9	1012	28	BZ563446	BZ563446 pac82-164
36	40.4	1.9	1201	13	BX399990	BX399990 BX399990
37	40.2	1.9	924	29	CNS03ACB	AL235460 Tetraodon
38	40	1.9	423	29	P132R	AL160522 Leishmani
39	40	1.9	983	28	BZ558308	BZ558308 pac81-60
40	39.8	1.9	530	10	BG023387	BG023387 cg40h09.x
41	39.4	1.9	694	28	AQ989725	AQ989725 Rfc00358
42	39.4	1.9	699	28	BZ560591	BZ560591 pac82-164
43	39.4	1.9	920	29	CNS0062R	AL061710 Drosophil
44	39.2	1.9	855	29	CNS011PE	AL100556 Drosophil
45	39	1.9	1201	13	BX361080	BX361080 BX361080

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:521013
 This read is a RESEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the correct orientation)
 Seq primer: -4ORP.

FEATURES

source Location/Qualifiers
 1. .290
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J x DBA/2J F1"
 /db_xref="taxon:10090"
 /clone="IMAGE:893053"
 /dev_stage="embryo (post-implantation)"
 /lab_host="DH10B"
 /clone_lib="Knowles Solter mouse inner cell mass"
 /notes="Vector: pBluescript SK-; Site1: XbaI; Site2: XhoI; Cloned unidirectionally from mRNA prepared from primitive streak embryonic tissue. Primer: Oligo dt. cDNAs were cloned into the XbaI/XhoI sites of pBluescript SK+ (Stratagene) using commercial linkers (NEB). Average insert size: 0.5 kb."

ORIGIN

Query Match 13.7%; Score 286.8; DB 9; Length 290;
 Best Local Similarity 99.3%; Pred. No. 1.9e-71;
 Matches 288; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 722 TTGACTGGAGGCGTCGGAGCAGATGTCTCTGTTTGTATGATACCCCGGCGAGC 781
 Db 1 TTGACTGGAGGCGTCGGAGCAGATGTCTCTGTTTGTATGATACCCCGGCGAGC 60
 QY 782 GTTATGATAACCGGGATGGGCAACTCGGAGTCTGACGGGGGATGATACCGACCTTCG 841
 Db 61 GTTATGATAACCGGGATGGGCAACTCGGAGTCTGACGGGGGATGATACCGACCTTCG 120
 QY 842 GCTATGACGACACAAATTTGAGTGGCTATGATCATCTTTTCCCTTCGGAACATGA 901
 Db 121 GCTATGACGACACAAATTTGAGTGGCTATGATCATCTTTTCCCTTCGGAACATGA 180
 QY 902 AATCGTATCTGACTGGAGCAGACAGAAAAATAAGGTCGTGAGCTTGTACGAGTGTAC 961
 Db 181 AATCGTATCTGACTGGAGCAGACAGAAAAATAAGGTCGTGAGCTTGTACGAGTGTAC 240
 QY 962 TGAAGCGGACAAATCGGGGCTTCGGGTCAGCCGCGGAGCTTAAGGAA 1011
 Db 241 TGAAGCGGACAAATCGGGGCTTCGGGTCAGCCGCGGAGCTTAAGGAA 290

RESULT 2

AI481927/c
 LOCUS AI481927 289 bp mRNA linear EST 09-MAR-1999
 DEFINITION vH86a12.x1 Knowles Solter mouse embryonic stem cell Mus musculus
 cDNA clone IMAGE:893854 3', mRNA sequence.
 AI481927
 ACCESSION AI481927.1 GI:4375153
 VERSION
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 289)
 REFERENCE
 AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pace,D., Harvey,N., Schurk,R., Ritter,B., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
 TITLE The WashU-NCI Mouse EST Project 1999
 JOURNAL Unpublished (1999)
 COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:521814
 This clone was previously sequenced on the 5' end only, this new data is from the 3' end
 Seq primer: Primer name ambiguous.

FEATURES

source Location/Qualifiers
 1. .289
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J x DBA/2J F1"
 /db_xref="taxon:10090"
 /clone="IMAGE:893854"
 /dev_stage="embryo"
 /lab_host="DH10B"
 /clone_lib="Knowles Solter mouse embryonic stem cell"
 /notes="Vector: pSPORT; Site1: NotI; Site2: SalI; Cloned unidirectionally from mRNA prepared from 800 blastocysts. Primer: SalI(dT): 5'-CGTCACCGTCGACCGTTTTTTTTTTT-3'. cDNAs were cloned into the NotI/SalI sites of a pSPORT vector (Life Technologies)."

ORIGIN

Query Match 13.3%; Score 278; DB 9; Length 289;
 Best Local Similarity 98.3%; Pred. No. 6.7e-69;
 Matches 281; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 775 CGGCAGGTTTATGATAACCGGATGGGCAACTCGGAGTCTGACGGGGGATATGACCG 834
 Db 289 CGGCAGGTTTATGATAACCGGATGGGCAACTCGGAGTCTGACGGGGGATATGACCG 230
 QY 835 ACCTCGGCTATGAGCGAAACAAATTTGAGTGGCTATGATCATCTTTTCCCTTCGGA 894
 Db 229 ACCTCGGCTATGAGCGAAACAAATTTGAGTGGCTATGATCATCTTTTCCCTTCGGA 170
 QY 895 ACATGGAATCGTATCTGAACTGGAACGACAGAAAAATAAGGTCGTGAGCTTGTACCG 954
 Db 169 ACATGGAATCGTATCTGAACTGGAACGACAGAAAAATAAGGTCGTGAGCTTGTACCG 110
 QY 955 AGTGTACTGAAGCGGACAAATCGGGGCTTCGGGTCAGCCGCGGAGCTTAAGGAATCG 1014
 Db 109 AGTGTACTGAAGCGGACAAATCGGGGCTTCGGGTCAGCCGCGGAGCTTAAGGAATCG 50
 QY 1015 AACCTTATCTGAATTCATTACTGCTTACCCCTCTCGGAGATCTC 1060
 Db 49 AACCTTATCTGAATTCATTACTGCTTACCCCTCTCGGAGATCTC 4

RESULT 3

AZ935601/c
 LOCUS AZ935601 495 bp DNA linear GSS 16-SEP-2002
 DEFINITION T034 E. coli C1845 library subtracted from E. coli MG1655
 Escherichia coli genomic clone T034, genomic survey sequence.
 AZ935601
 ACCESSION AZ935601.1 GI:14579858
 VERSION
 KEYWORDS GSS.
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
 1 (bases 1 to 495)
 REFERENCE
 AUTHORS Blanc-Potard,A., Tinsley,C., Scaletsky,I., Le Bouguenec,C., Guignot,J., Servin,A.L., Nassif,X. and Bernet-Camard,M.
 TITLE Representational Difference Analysis between Afa/Dr Diffusely Adhering Escherichia coli and Nonpathogenic E. coli K-12
 JOURNAL Infect. Immun. 70 (10), 5503-5511 (2002)
 COMMENT Contact: Blanc-Potard A
 Inserm U510
 Faculte de Pharmacie Paris XI
 92296, Chatenay-Malabry cedex, France

Tel: 33 1 46 83 58 43
Fax: 33 1 46 83 58 44

Email: anne.blanco@pseud.fr

Class: representational difference analysis.

Location/Qualifiers

1. 495

/organism="Escherichia coli"

/mol_type="genomic DNA"

/strain="C1845"

/db_xref="taxon:562"

/clone="T034"

/clone_lib="E. coli C1845 library subtracted from E. coli

MG1655"

ORIGIN

Query Match 12.0%; Score 251; DB 28; Length 495;
Best Local Similarity 90.0%; Pred. No. 6.2e-61;
Matches 280; Conservative 0; Mismatches 30; Indels 1; Gaps 1;
QY 1753 GCATCATGGCTGGTGGCCGTTATCGCGGGGAAAACACCCAGCTTTTACCCAGAAATTATTCG 1812
DB 311 CGGTCGTACCTGGCGGGGTTTCGACGGGTTACAAACCGTGGCCAGCGTGCGAATTATTCG 252
QY 1813 TCATGACCGCTGTACAGAAAGTGTATGATGAGAAAGAGATACCTGAAAGCCTGG 1872
DB 251 TCACTGACCGCTGTACAGAAAGTGTATGATGAGAAAGAGATACCTGAAAGCCTGG 192
QY 1873 ACGGTGGTGGATGACGGTCTGTCTGGGAAGATGACGGATGCCCTGACCTGAATGCTGGC 1932
DB 191 ACGGTGGTGGATGACGGTCTGTCTGGGAAGATGACGGATGCCCTGACCTGAATGCTGGC 132
QY 1933 GTGAATAACCT-GCTCAACAGGATTACAGTACGCTGACGCTGTACAGTCCGCTAAGAG 1991
DB 131 GTGAATAACCTAGCTCAACAGGATTACAGTACGCTGACGCTGTACAGTCCGCTAAGAG 72
QY 1922 TACGCTGTATCCCGTGATTCTCCACACGGATCATCAACAGGATGATGATACC 2051
DB 71 TACGCTGTATCCCGTGATTCTCCACACGGATCATCAACAGGATGATGATACC 12
QY 2052 TGAGCGAAATT 2062
DB 11 TGAGCGAAATT 1

RESULT 4
AA501138
LOCUS
DEFINITION
vh77h07.r1 Knowles Solter mouse inner cell mass Mus musculus cDNA
clone IMAGE:893053 5', mRNA sequence.
AA501138
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 243)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.

MG1:521013.

Location/Qualifiers

1. 243

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J x DBA/2J F1"

/db_xref="taxon:10090"

/clone="IMAGE:893053"

/dev_stage="embryo (post-implantation)"

/lab_host="DH10B"

/clone_lib="Knowles Solter mouse inner cell mass"

/note="Vector: pBluescript SK+; Site 1: XbaI; Site 2:

XhoI; Cloned unidirectionally from mRNA prepared from

primitive streak embryonic tissue. Primer: Oligo dr.

CDNAs were cloned into the XbaI/XhoI sites of pBluescript

SK+ (Stratagene) using commercial linkers (NEB). Average

insert size: 0.5 kb."

ORIGIN

Query Match 10.7%; Score 224.6; DB 9; Length 243;
Best Local Similarity 97.9%; Pred. No. 1.8e-53;
Matches 238; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 803 AACTGGGGAGTCTGACGGGGGATATGACCGACCTCGGCTATGAGCGAA-ACAAAATT 861
DB 1 AACTGGGGAGTCTGACGGGGGATATGACCGACCTCGGCTATGAGCGAA-ACAAAATT 60
QY 862 TCAGCTGGCTGTATGATCATATTTTCACTTCGGAAACATGGAATCGTATCTGAACCTGGAAC 921
DB 61 TCAGCTGGCTGTATGATCATATTTTCACTTCGGAAACATGGAATCGTATCTGAACCTGGAAC 120
QY 922 GACACAGAAAATAAGGTGTGAGCTGTACCGAGTGTACTGAAGCGCGACAAATGGGG 981
DB 121 GACACAGAAAATAAGGTGTGAGCTGTACCGAGTGTACTGAAGCGCGACAAATGGGG 180
QY 982 CTTGCCCGCTCAGCGCGGGAGCTTAAAGGAATCGAACCTTATCCTGAATTCATTACTGCTT 1041
DB 181 CTTGCCCGCTCAGCGCGGGAGCTTAAAGGAATCGAACCTTATCCTGAATTCATTACTGCTT 240
QY 1042 ACC 1044
DB 241 ACC 243

RESULT 5

AQ990279
LOCUS
DEFINITION
AQ990279.1 GI:9648873
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Photorhabdus luminescens
Photorhabdus luminescens
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Photorhabdus.
1 (bases 1 to 586)
ffrench-Constant, R.H., Waterfield, N., Burland, V., Perna, N.T.,
Daborn, P.J., Bowen, D. and Blattner, F.R.
A genomic sample sequence of the entomopathogenic bacterium
Photorhabdus luminescens W14: potential implications for virulence
Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: ffrench-Constant RH

Department of Biology and Biochemistry

University of Bath

South Building, Bath BA2 7AY, UK

Tel: (44) 1225 826621

Fax: (44) 1225 826779

Email: bsrfr@bath.ac.uk

This is one of 2,122 random reads from the M13 library. For

annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see french-Constant et al. 2000, Nucleic Acids Res.

Seq primer: M13 Forward
Class: shotgun.

FEATURES

source Location/Qualifiers

1..586

/organism="Photorhabdus luminescens"

/mol_type="genomic DNA"

/strain="W14"

/db_xref="taxon:29488"

/clone="PLG01023"

/dev_stage="primary phase variant"

/clone_lib="Photorhabdus luminescens strain W14 M13 library"

/notes="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus."

ORIGIN

Query Match 3.9%; Score 81.8; DB 28; Length 586;

Best Local Similarity 53.4%; Pred. No. 6.8e-12; Indels 6; Gaps 1;

Matches 198; Conservative 0; Mismatches 167;

1150 AGCTGCTGGTATTGCTGAGGATGAGTGGCATCTCACGGATGCATTTGCGCTGACTGCG 1209

133 AGTTGGCGTGTGTTCTGAGGATGAGTGGCATGACCAATGACTTTTCTTGGTGGT 192

1210 GGCAGCGGTATGAACATCATGAGCAATTCGGGGACACATTCAGTCCGGTGCATATCTG 1269

193 GGTATCCGTATGATTAAGACGAGAAATTTGGCAGTCATTTGCACTCCACGCTTATGTT 252

1270 GTCTGGGATGTGGCAGATGCTGGACCTGAAAGCGGTGACCAACGGATATAAGCA 1329

253 GTGTGCATGTAGATGAACATGGCAATTAAGGTGGTGTATCTACAGTTACCGTTCT 312

1330 CCAGAAATGGGCA-----GCTACATAAAGGATAGTGGTGTGTCCTGGGAGGAAA 1383

313 CCGGATTTGCTTATATTTTCGGCGGATTTGGGGGAGCAAAACAGGTGTTATGTTGGAAAT 372

1384 ACAATCTACTTGTGAACCCGACCTGAAGCGGAGAGAGCGTCACTGTTAGGCTGGG 1443

373 GGTATTATTTCTGGTAACCGAATTTGAACACAGAGAAAGTGTCAAGAGAAATTAAT 432

1444 GTGTAATACCAATACCCCGCTGTAATGCCAATGTACAGGTTTATGACTGACTTC 1503

433 GTATGTGGAACATCCGGATAATTTAATGCTGGATTGACCATATTTAATACTGATTT 492

1504 TCCAAACAGAT 1514

493 AAGATAAAAT 503

RESULT 6

CF199793/c

LOCUS

DEFINITION

EST1397 Tamarix androssowii leaf Tamarix androssowii cDNA, mRNA

sequence.

ACCESSION

CF199793

VERSION

CF199793.1 GI:33394166

KEYWORDS

EST.

SOURCE

Tamarix androssowii

ORGANISM

Tamarix androssowii

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Caryophyllales; Tamaricaceae; Tamarix.

1 (bases 1 to 690)

Wang, Y., Yang, C., Jiang, J., Liu, G., Wu, J. and Liu, Z.

EST acquired from cDNA library of Tamarix androssowii treated with

NaHCO3

Unpublished (2003)

JOURNAL

CONTACT: Yucheng Wang

Forestry Source and Environment College

Northeast Forestry University

FEATURES

source

1..988

/organism="Pseudomonas aeruginosa"

/mol_type="genomic DNA"

/strain="WSH"

/db_xref="taxon:287"

/clone="msh2_5047"

/clone_lib="msh"

/notes="Environmental isolate. Whole genomic shotgun library."

ORIGIN

Query Match 3.2%; Score 67.4; DB 14; Length 690;

Best Local Similarity 53.6%; Pred. No. 1.1e-07; Indels 0; Gaps 0;

Matches 140; Conservative 0; Mismatches 121;

388 ATTGAGCGTATTGAGGTTATCAGGGGCGGATGTCACACTGTATGGCTCTGATGCGATG 447

302 ATTGACGCTATTGAAGTTCTGCGTGGTCCGCGAGCTGCGGTTATGCAACGCGCGCG 243

448 GCGGCTGCTGTAATCATTTACAGAAAGATGCAGCAAAATGGCTCTCTCCGTCAT 507

242 GCGGCGGTGTTAACTATCATTTACAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 183

508 GCAGGGCTGAATCTGAGGAAAGCAAAATGGGGTAACAGCAGCCAGTTTAAATTTCTGG 567

182 GCATATTTCAATCGCCAGACATAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 123

568 AGCAGTGGTCCCTTTGGGATGATTTCTGTCAGCTCGAGTACCGGTCAGCAGCAGCAG 627

122 CTGACCGGTCGCTGGGCGACGAATTCGACAGGCTTGTATCAATTTGTGCTCTATAC 63

628 CGTCAGGGTTTCATCGGTCACA 648

62 TAAACAAGTGTAACTGTACA 42

Hexing 26, Harbin, Heilongjiang, 150040, P.R. China
Tel: 086-451-2190607
Email: WANGYUCHENG1029@YAHOO.COM.CN.

FEATURES

source

1..690

/organism="Tamarix androssowii"

/mol_type="mRNA"

/db_xref="taxon:189785"

/tissue_type="leaf"

/clone_lib="Tamarix androssowii leaf"

ORIGIN

Query Match

Best Local Similarity

Matches

388

302

448

242

508

182

568

122

628

62

RESULT 7

BZ576672/c

LOCUS

DEFINITION

msh2_5047.y2 msh Pseudomonas aeruginosa genomic clone msh2_5047,

genomic survey sequence.

ACCESSION

BZ576672

VERSION

BZ576672.1 GI:27211733

KEYWORDS

GSS.

SOURCE

ORGANISM

Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

1 (bases 1 to 988)

Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

Burns, J.L., Kaul, R. and Olsen, M.V.

Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol. (2002) In press

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352445, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun

Location/Qualifiers

1..988

/organism="Pseudomonas aeruginosa"

/mol_type="genomic DNA"

/strain="WSH"

/db_xref="taxon:287"

/clone="msh2_5047"

/clone_lib="msh"

/notes="Environmental isolate. Whole genomic shotgun library."

FEATURES

source

1..988

/organism="Pseudomonas aeruginosa"

/mol_type="genomic DNA"

/strain="WSH"

/db_xref="taxon:287"

/clone="msh2_5047"

/clone_lib="msh"

/notes="Environmental isolate. Whole genomic shotgun library."

ORIGIN

Query Match

Best Local Similarity

Matches

388

302

448

242

508

182

568

122

628

62


```

ORIGIN
Query Match          3.1%; Score 64.8; DB 28; Length 988;
Best Local Similarity 55.9%; Pred. No. 7.9e-07;
Matches 123; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 1199 CGCTGACTGCGGCGAGCGCTATGAACATCATGATGAGCAATTCGGGGGACACTTCAGTCCGC 1258
      |||
Db 392 CGCTGACCTCGGAGAGCGCTTCACACAGCAGAGGGTTCGGTGGGAAGCAGTCCGC 333

QY 1259 GTGCATATCTGCTCTGGGATGTGCAGATGCTCGAGCGCTGAAAGCGGTGTGACACAGG 1318
      |||
Db 332 GCCTGTACTCTGCTGACACACCTTTCCGATGCGCTGACCTCGCGCGCGGTGCGGCGCG 273

QY 1319 GATATAGCCACCCAGATGGGAGCTACATAAGGGATTAGTGGTGTGTCGGGCGAGG 1378
      |||
Db 272 GCTCAAGGCGCCGAGCCTGAAGCAGCTATCGCGGAGTATCGCGGCTGGTGGCGCG 213

QY 1379 GAAAAACAAATCTACTTGTGTAACCCCGACCTGAAAGCCGGA 1418
      |||
Db 212 GCGGCTTCCACATCTACGGCAATCCGACCTGAAAGCCGGA 173

RESULT 8
CG745412/c
LOCUS
DEFINITION
P038-2-C09.za Ppa EcoRI BAC library Pristionchus pacificus genomic,
genomic survey sequence.
ACCESSION
CG745412
VERSION
CG745412.1 GI:37966338
KEYWORDS
GSS.
SOURCE
Pristionchus pacificus
ORGANISM
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE
1 (bases 1 to 1350)
Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Buntjer,J., van der Meulen,M. and Sommer,R.J.
An integrated physical and genetic map of the nematode Pristionchus
pacificus
Mol. Genet. Genomics 269 (5), 715-722 (2003)
22835951
12884007
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spenannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: raif.sommer@tuebingen.mpg.de
Class: BAC ends.
Location/Qualifiers
1..1350
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector."
FEATURES
source

ORIGIN
Query Match          3.0%; Score 62.4; DB 29; Length 1350;
Best Local Similarity 57.0%; Pred. No. 4.7e-06;
Matches 114; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 388 ATTGAGCGTATTGAGTTATCAGGGCGGATGTCCACACTGTATGCTCTGATCGATG 447
      |||
Db 205 ATTGACGATTGAGTTCTGCTGGTCCGGAGTTCGGCTTATGGCAAGCGCGCGG 146

QY 448 GCGGTGTGGTGAATATATATACCAAGAAATGACAGAAATGCGTCTCTTCCTCAAT 507
      |||

```

```

Db 145 GGCGCGTGGTTAACATCATATTACCAAAAAAGCAGCGGAGTGGCAGCGCTCTCTGGAC 86
QY 508 GCAGGCGCTGAATCTGCAGGARRAGCAACAAATGGGTAAACAGCAGCCAGTTTAAATTCGG 567
      |||
Db 85 GCATATTTCAATGGCGCAGACATAAAGAGGAGGTGCCCAACAGGCACTTAACITTAGC 26
      |||
QY 568 AGCAGTGTGTCCTCTTGTGGA 587
      |||
Db 25 CTGACCGGTCCGCTGGCGA 6

RESULT 9
AZ302660/c
LOCUS
DEFINITION
GSSBru1597 Brucella abortus random genomic library Brucella
melitensis biovar Abortus genomic clone UU1597, genomic survey
sequence.
ACCESSION
AZ302660
VERSION
AZ302660.1 GI:10128871
KEYWORDS
GSS.
SOURCE
Brucella melitensis biovar Abortus (Brucella abortus)
ORGANISM
Brucella melitensis biovar Abortus
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Brucellaceae; Brucella.
REFERENCE
1 (bases 1 to 519)
Sanchez,D.O., Zandomeni,R.O., Cravero,S., Verdun,R.E., Pierrou,E.,
Faccio,P., Diaz,G., Lanzavecchia,S., Aguero,F., Frasch,A.C.C.,
Andersson,S.G.E., Rosetti,O.L., Grau,O. and Ugalde,R.A.
Gene discovery through genomic sequencing of Brucella abortus
Infect. Immun. 69 (2), 865-868 (2001)
21101034
11159979
Contact: Siv Andersson
Small Genomes Sequencing Group
Department of Molecular Evolution, Uppsala University
Norbyvagen 18C, S-752 36, Uppsala, Sweden
Tel: 46-18-471-4379
Fax: 46-18-471-6404
Email: Siv.Andersson@ebc.uu.se
Sequences were basecalled with phred and vector was masked with
crossmatch (see http://genome.washington.edu). Sequences were then
trimmed from both ends to remove low quality bases and masked
vector.
Class: shotgun.
Location/Qualifiers
1..519
/organism="Brucella melitensis biovar Abortus"
/mol_type="genomic DNA"
/strain="2308"
/db_xref="taxon:235"
/clone_lib="UU1597"
/clone_lib="Brucella abortus random genomic library"
/note="Vector: modified M13"
FEATURES
source

ORIGIN
Query Match          2.7%; Score 57.2; DB 28; Length 519;
Best Local Similarity 72.5%; Pred. No. 8.4e-05;
Matches 74; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 376 CCTCTGCGCCCGCATTTAGCGTATTTAGGTTATCAGGGGCGGATGTCACACTGTATGGC 435
      |||
Db 354 CGCTGACCTCCATCGAACGTTATCGAGATTGCGGGGCGGCGATTCTCTCGCAATATGGC 295

QY 436 TCTGATCGATGGCGGTGTGGTGAATATATCATACCAGAAAG 477
      |||
Db 294 GCGATGCAATCGCGGTGTGTCATCATATATCATCCCAACAG 253

RESULT 10
BZ568287
LOCUS
DEFINITION
BZ568287 1029 bp DNA linear GSS 17-DEC-2002
pacs2-164_7457.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
pacs2-164_7457, Genomic survey sequence.

```


FEATURES
source

processed analyzed, groomed and submitted by the Animal Biotechnology Center (URL: <http://primer.ansci.umn.edu/software.html>), University of Minnesota. Trace files have been submitted to the NCBI trace archive (<http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?>). Chromatograms were analyzed with Phred Version: 0.000925.c. Vector (pCMV SPORT6, Invitrogen), Sus scrofa mitochondrion (gi|5835862|ref|NC_000845.1), Escherichia coli K12 (gi|5835862|ref|NC_000913.1), Forcine reproductive and respiratory syndrome virus (gi|11878202|gb|AF303357.1|AF303357), Altermonas sp.16S rRNA gene (gi|4218471|emb|Y15322.1|PLX15322), Sus scrofa 18S ribosomal RNA (gi|3388170|gb|AF080393.1|AF080393), and Sus scrofa 18S ribosomal RNA (gi|3388592|gb|AF102857.1|AF102857). Sequences were identified and masked by Crossmatch -minscore=20, -mismatch=12, -penalty=-8.

SWAR alignment scores were set to match score=1, gap initial penalty=-3 and gap extension penalty=-2. A continuous stretch of at least 50 bases of a PHRED quality 20 were required for submission.5' and 3' TERMINI were defined by a continuous stretch of 10 nucleotides with a PHRED quality not less than 20. A terminal stretch of at most 40 un-masked bases were automatically trimmed when flanking masked sequence. Sequences with an INTERNAL continuous stretch of at most 20 bases with PHRED quality less than 20 were automatically prepared for submission. HIQUAL_START and HIQUAL_STOP refer to the coordinates on CLIP_QUALITY and CLIP_VECTOR at the NCBI trace archive.

Plate: 29 row: G column: 05

Seq primer: SP6 primer
High quality sequence start: 58
High quality sequence stop: 245.

FEATURES

source

1. 609
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/clone="PPSUBLIB_29G05"
/tissue_type="Peyer's patch"
/clone_lib="UMMPM3"
/note="Organ: small intestine, jejunum; Vector: pCMVSPORT6; Site 1: NotI; Site 2: SalI; Jejunal Peyer's patches were isolated from healthy, 4-6 week old cross-bred pigs. RNA was extracted either immediately or after 3 hours stimulation in an Using chamber with one of the following treatments: 1) no treatment, 2) Salmonella choleraesuis vaccine strain SC-54, 3) phorbol myristate acetate, concanavalin A, and 8-bromo-cyclic AMP, and 4) lipopolysaccharide and cholera toxin. Each treatment was performed in the presence and absence of cycloheximide. Purified poly A + RNA from each of the treated tissues (2-4) was combined together, reverse transcribed, and cloned in to pCMVSPORT6 to make a library of approximately 530,000 recombinant clones with an average insert size of 1.0 kbp. Poly A+ RNA from freshly isolated and non-treated tissue from an Using chamber (treatment 1) was cloned in the same manner to produce an unstimulated cDNA library of approximately 900,000 clones with an average insert size of 1.5 kbp. Equal portions of the two libraries were pooled and then subtracted with porcine ST fibroblast RNA to create a subtracted porcine Peyer's patch library of approximately 6,000 clones with an average insert size of 1.0 kbp."

ORIGIN

Query Match 2.5%; Score 51.6; DB 14; Length 609;
Best Local Similarity 70.4%; Pred. No. 0.0039;
Matches 69; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
375 CCTCTGGCGCATTCAGGATTCAGGTTATCAGGGGCGGATCCACATGATGG 434
177 CCTATTGGCTTGTCCAGCGTTGAAATATATCCGTGGCGCGCTCCGCTTTATGG 236
435 CTCTGATGCATGGCGGGTGTGGTGAATATCAFTACCA 472
237 TTCGATGCAATAGCGGGGTGTGATATCATCATCAGCA 274

RESULT 15
CB286521
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CB286521 618 bp mRNA linear EST 27-FEB-2003
CMD39_B07_60 UMMPM3 Sus scrofa cDNA clone PPSUBLIB_24B07 5', mRNA
sequence.
CB286521 GI:28577475
EST.
Sus scrofa (pig)
Sus scrofa
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 618)
Dvorak,C.M.T., Hyland,K.A., Zhang,Y., Fahrenkrug,S.C. and
Murtaugh,M.P.
Porcine jejunal Peyer's patch expressed sequences
Unpublished (2003)
Contact: Murtaugh, M.P.
Department of Veterinary Pathobiology
University of Minnesota
1971 Commonwealth Ave., St. Paul, MN 55108, USA
Tel: 612-625-6735
Fax: 612-625-5203
Email: murta001@umn.edu

The PHRED quality scores of this sequence are supplied below: 39 39
39 25 22 22 22 22 31 46 51 51 56 56 56 56 56 56 56 56 56 56
56 56 51 51 51 51 51 51 56 56 56 51 51 45 34 32 32 34 34 35
35 40 51 51 51 51 51 51 56 56 35 35 35 29 29 29 29 31 46
35 35 38 38 35 35 40 40 40 40 40 40 40 40 40 40 40 40
46 43 43 45 43 35 35 39 32 32 32 35 35 56 56 56 56
56 35 35 35 35 35 42 46 46 56 45 43 43 43 43 43 43 43 43 43
45 51 51 51 51 51 56 56 56 56 51 51 51 43 43 41 41 56 56
56 45 45 45 43 43 38 43 36 45 46 46 29 28 28 35 35 56 56
56 56 42 41 41 45 45 51 51 43 40 40 38 38 56 56 56 56
56 56 43 43 46 43 43 46 43 46 43 46 43 46 43 46 43 46 43
51 51 45 45 45 45 45 46 41 41 46 46 51 51 56 56 56 56 56 56
36 36 36 40 40 43 51 45 45 45 45 45 45 45 45 45 45 45 45
51 43 43 43 43 43 56 56 56 46 43 43 43 43 43 43 43 43 43 43
45 51 56 56 56 56 56 51 45 45 43 43 43 43 43 43 43 43 43 43
40 43 43 43 56 56 56 51 51 45 45 45 45 45 45 45 45 45 45 45
56 43 43 56 56 56 56 43 43 43 43 38 38 38 38 38 38 38 38 38 38
43 35 35 35 40 40 45 45 51 39 38 38 38 38 38 38 38 38 38 38
45 45 51 51 51 51 43 46 43 35 35 35 35 35 35 35 35 35 35 35
56 37 37 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
45 45 51 51 45 45 40 40 37 37 37 37 37 37 37 37 37 37 37 37
45 51 51 51 40 40 37 35 35 35 35 35 35 35 35 35 35 35 35 35
40 37 40 37 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
36 44 34 32 32 29 29 29 29 29 29 29 29 29 29 29 29 29 29 29
25 29 34 30 32 29 29 29 29 29 29 29 29 29 29 29 29 29 29 29
29 29 29 29 29 29 34 34 34 34 34 34 34 34 34 34 34 34 34 34
28 29 29 29 29 29 32 32 32 32 32 32 32 32 32 32 32 32 32 32
Sequences were generated on ABI 3100 and 3700 capillary sequencers.
Raw sequence data was batch processed analyzed, groomed and
submitted by the Animal Biotechnology Center (URL:
<http://primer.ansci.umn.edu/software.html>), University of
Minnesota. Trace files have been submitted to the NCBI trace
archive (<http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?>).
Chromatograms were analyzed with Phred Version: 0.000925.c. Vector
(pCMV SPORT6, Invitrogen), Sus scrofa mitochondrion
(gi|5835862|ref|NC_000845.1), Escherichia coli K12 (gi|5835862|
ref|NC_000913.1), Forcine reproductive and respiratory syndrome
virus (gi|11878202|gb|AF303357.1|AF303357), Altermonas sp.16S rRNA
gene (gi|4218471|emb|Y15322.1|PLX15322), Sus scrofa 18S ribosomal
RNA (gi|3388170|gb|AF080393.1|AF080393), and Sus scrofa 18S ribosomal
RNA (gi|3388592|gb|AF102857.1|AF102857). Sequences were identified
and masked by Crossmatch -minscore=20, -mismatch=12, -penalty=-8.
SWAR alignment scores were set to match score=1, gap initial
penalty=-3 and gap extension penalty=-2. A continuous stretch of at
least 50 bases of a PHRED quality 20 were required for

submission.5' and 3' TERMINI were defined by a continuous stretch of 10 nucleotides with a PHRED quality not less than 20. A terminal stretch of at most 40 un-masked bases were automatically trimmed when flanking masked sequences. Sequences with an INTERNAL continuous stretch of at most 20 bases with PHRED quality less than 20 were automatically prepared for submission. HIQUAL START and HIQUAL STOP refer to the coordinates on CLIP_QUALITY and CLIP_VECTOR at the NCBI trace archive.

Plate: 24 row: B column: 07
Seq primer: SP6 primer
High quality sequence start: 57
High quality sequence stop: 265.

FEATURES

source

1. .618
Location/Qualifiers
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/clone="PPSUBLIB_24B07"
/tissue_type="Peyer's patch"
/clone_lib="UMNMPM3"

/note="Organ: small intestine, jejunum; Vector: pCMVSPORT6; Site 1: NotI; Site 2: SalI; Jejunal Peyer's patches were isolated from healthy, 4-6 week old cross-bred pigs. RNA was extracted either immediately or after 3 hours stimulation in an Ussing chamber with one of the following treatments: 1) no treatment, 2) Salmonella choleraesuis vaccine strain SC-54, 3) phorbol myristate acetate, concanavalin A, and 8-bromo-cyclic AMP, and 4) lipopolysaccharide and cholera toxin. Each treatment was performed in the presence and absence of cycloheximide. Purified poly A + RNA from each of the treated tissues (2-4) was combined together, reverse transcribed, and cloned in to pCMVSPORT6 to make a library of approximately 530,000 recombinant clones with an average insert size of 1.0 kbp. Poly A+ RNA from freshly isolated and non-treated tissue from an Ussing chamber (treatment 1) was cloned in the same manner to produce an unstimulated cDNA library of approximately 900,000 clones with an average insert size of 1.5 kbp. Equal portions of the two libraries were pooled and then subtracted with porcine ST fibroblast RNA to create a subtracted porcine Peyer's patch library of approximately 6,000 clones with an average insert size of 1.0 kbp."

ORIGIN

Query Match 2.5%; Score 51.6; DB 14; Length 618;
Best Local Similarity 70.4%; Pred. No. 0.0039;
Matches 69; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY	375	CCCTCTGGCGCCATTGACGGTATTGAGTTATCAGGGGGCGGATGCCACACTGTATGG	434
Db	177	CCCTATTGGCTGTGCCAGCGTGTGATATATCCGTGGCGCGCTCCGCTGTTATGG	236
QY	435	CTCTGATGCGATGGCGGTGTGGTGAATATCATATTACCA	472
Db	237	TTCGGATGATAGCGGGGTGGTGAATATCATACGA	274

Search completed: October 12, 2004, 13:40:42
Job time : 5412 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2004, 13:40:48 ; Search time 999 Seconds
(without alignments)

10611.075 Million cell updates/sec

Title: US-10-625-972-4

Perfect score: 2091

Sequence: 1 atgcgaataaccactctggc.....cgctgaactatcagttctga 2091

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3340653 seqs, 2534783454 residues

Total number of hits satisfying chosen parameters: 6681306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	2089.4	99.9	87563	15	US-10-114-170-57
C 2	2079.8	99.5	76804	15	US-10-085-959-109
C 3	125.6	6.0	2046	15	US-10-238-075-1076
C 4	125.6	6.0	12943	15	US-10-238-075-1049
C 5	64.2	3.1	758	11	US-09-764-875-112
C 6	64.2	3.1	822	9	US-09-764-868-94
C 7	57.6	2.8	686	9	US-09-764-868-510
C 8	57.6	2.8	686	11	US-09-764-875-424
C 9	51.8	2.5	296	11	US-09-864-408A-851
C 10	51.4	2.5	2731748	17	US-10-238-075-526
C 11	48.6	2.3	1362	15	US-10-238-075-507
C 12	48.6	2.3	12264	15	US-10-085-959-11
C 13	48.6	2.3	23654	15	US-10-085-959-11
C 14	47	2.2	2760	13	US-10-194-163-462

15	43	2.1	2061	13	US-10-282-122A-12338	Sequence 12338, A
16	42	2.0	2182	13	US-10-194-163-916	Sequence 916, App
17	39.6	1.9	4534	13	US-10-194-163-619	Sequence 619, App
18	39.4	1.9	990	13	US-10-282-122A-12119	Sequence 12119, A
19	39.2	1.9	699	15	US-10-184-644-138	Sequence 138, App
20	39.2	1.9	699	15	US-10-184-634-138	Sequence 389, App
21	38.8	1.9	2127	16	US-10-264-237-389	Sequence 126, App
22	38.8	1.9	4758	15	US-10-175-523-126	Sequence 12731, A
23	38.4	1.8	2052	13	US-10-282-122A-12731	Sequence 12007, A
24	38	1.8	576	15	US-10-029-386-12007	Sequence 24900, A
25	38	1.8	674	15	US-10-029-386-20342	Sequence 20342, A
26	37.6	1.8	1184	13	US-10-142-426-394	Sequence 394, App
27	37.6	1.8	1184	15	US-10-123-155-394	Sequence 394, App
28	37.6	1.8	1184	15	US-10-146-731-394	Sequence 394, App
29	37.6	1.8	1184	15	US-10-140-472-394	Sequence 394, App
30	37.6	1.8	1184	15	US-10-141-761-394	Sequence 394, App
31	37.6	1.8	1184	15	US-10-142-885-394	Sequence 394, App
32	37.6	1.8	1184	15	US-10-158-790-394	Sequence 394, App
33	37.6	1.8	1184	16	US-10-137-871-394	Sequence 394, App
34	37.6	1.8	1184	16	US-10-140-923-394	Sequence 394, App
35	37.6	1.8	1184	16	US-10-141-759-394	Sequence 394, App
36	37.6	1.8	1184	16	US-10-141-759-394	Sequence 394, App
37	37.6	1.8	1184	16	US-10-141-759-394	Sequence 394, App
38	37.6	1.8	1184	16	US-10-140-805-394	Sequence 394, App
39	37.6	1.8	1184	16	US-10-140-864-394	Sequence 394, App
40	37.2	1.8	2139	15	US-10-238-075-262	Sequence 262, App
41	37.2	1.8	2364	13	US-10-282-122A-22619	Sequence 22619, A
42	37.2	1.8	2890	9	US-09-881-752A-361	Sequence 361, App
43	37.2	1.8	7306	15	US-10-085-959-152	Sequence 152, App
44	37.2	1.8	7315	15	US-10-238-075-261	Sequence 261, App
45	36.8	1.8	300	11	US-09-864-408A-5399	Sequence 5399, App

ALIGNMENTS

RESULT 1

US-10-114-170-57/c
; Sequence 57, Application US/10114170
; Publication NO. US20030023075A1
; GENERAL INFORMATION:

APPLICANT: Blattner, Frederick R.

Perna, Nicole T.

Plunkett, Guy

Welch, Rod

TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli 0157

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/114,170

FILING DATE: 01-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/453,702

FILING DATE: 03-DEC-1999

APPLICATION NUMBER: 60/110,955

FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960296.95017


```
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 87563
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 57:
US-10-114-170-57

Query Match          99.9%; Score 2089.4; DB 15; Length 87563;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2090; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCGAATAACCACTCTGGCTTCCGTAGTCAATCCCTGTCTCGGATTTTCAGCCAGCAGC 60
DB 49991 ATGCGAATAACCACTCTGGCTTCCGTAGTCAATCCCTGTCTCGGATTTTCAGCCAGCAGC 49932

QY 61 ATAGCTGCTGCAGAGATGTGATGATTGCTCGGCATCCGCTATCAGAAAAGCTGACT 120
DB 49931 ATAGCTGCTGCAGAGATGTGATGATTGCTCGGCATCCGCTATCAGAAAAGCTGACT 49872

QY 121 AACCGAGCCCGCAGTGTCTGTGATTAGCCAGGAGGAATTGCAGTCCAGCCAGTACCAC 180
DB 49871 AACCGAGCCCGCAGTGTCTGTGATTAGCCAGGAGGAATTGCAGTCCAGCCAGTACCAC 49812

QY 181 GATCTGGCGGAGGCTCTGAGATCAGTAGAGGTGTGGATGTTGAAAGTGTGACGGTAA 240
DB 49811 GATCTGGCGGAGGCTCTGAGATCAGTAGAGGTGTGGATGTTGAAAGTGTGACGGTAA 49752

QY 241 ACCGAGGCTCGAATCAGCATCCAGGAATCCAGCAATCCAGCAATCCAGCAATCCAGCAAT 300
DB 49751 ACCGAGGCTCGAATCAGCATCCAGGAATCCAGCAATCCAGCAATCCAGCAATCCAGCAAT 49692

QY 301 GATGTGTTTCTGAGGCGGAGAGTGAAGTGAATCCCAACGGTCTTCTGCGATGAAT 360
DB 49691 GATGTGTTTCTGAGGCGGAGAGTGAAGTGAATCCCAACGGTCTTCTGCGATGAAT 49632

QY 361 ACCGGGTTTCATCCCTCTGCGCCCATTTGAGCTGATGAGTGTATCAGGGGCGCGATG 420
DB 49631 ACCGGGTTTCATCCCTCTGCGCCCATTTGAGCTGATGAGTGTATCAGGGGCGCGATG 49572

QY 421 TCCACACTGTATGGCTCTGATCGATGGCGGTGTGGTGAATATCATTTACCAGAAAGAA 480
DB 49571 TCCACACTGTATGGCTCTGATCGATGGCGGTGTGGTGAATATCATTTACCAGAAAGAA 49512

QY 481 GCAGCAATATGCTCTCTTCGTCATTCAGGGCTGATTCAGGAGGAGCAACAAATGG 540
DB 49511 GCAGCAATATGCTCTCTTCGTCATTCAGGGCTGATTCAGGAGGAGCAACAAATGG 49452

QY 541 GGTAAACAGCAGCCAGTTAAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTCTGTGAGC 600
DB 49451 GGTAAACAGCAGCCAGTTAAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTCTGTGAGC 49392

QY 601 CTGCGAGGTACCGGTAGCAGCAACAGCGGTGAGGTTTATCGGTCAATCATCTGAGCGAT 660
DB 49391 CTGCGAGGTACCGGTAGCAGCAACAGCGGTGAGGTTTATCGGTCAATCATCTGAGCGAT 49332

QY 661 ACAGCAGGCGCAGTATTCCTTATCCACGGAGTGCAGAGAATTATAATCTTGTGGTCACT 720
DB 49331 ACAGCAGGCGCAGTATTCCTTATCCACGGAGTGCAGAGAATTATAATCTTGTGGTCACT 49272

QY 721 CTTGACTGGAAGGCTCGAGCAGGATGTCTCTGGTTTGTATGATATCATCCCGGAG 780
DB 49271 CTTGACTGGAAGGCTCGAGCAGGATGTCTCTGGTTTGTATGATATCATCCCGGAG 49212

QY 781 CGTTATGATAACCGGATGGGCAACTGGGGAGTCTGACGGGGGATATGACCGGACCTTG 840
DB 49211 CGTTATGATAACCGGATGGGCAACTGGGGAGTCTGACGGGGGATATGACCGGACCTTG 49152
```


Db 48071 CTGAATGCTGGGTAATAACCTGCTCAACAAGATTACAGTACGAGCCCTGTACGT 48012
Qy 1981 GCCGGTAAGAGTACGCTGTATGCGGGTGATTAATTCAGACGGGATCAATCAACAACAGGA 2040
Db 48011 GCCGGTAAGAGTACGCTGTATGCGGGTGATTAATTCAGACGGGATCAATCAACAACAGGA 47952
Qy 2041 TATGTGATCTAGCGCAATTAATCTGATGTCGCTGAATCAATCAATCAATCAATCAATCA 2091
Db 47951 TATGTGATCTAGCGCAATTAATCTGATGTCGCTGAATCAATCAATCAATCAATCAATCA 47901

RESULT 2

US-10-085-959-109/c
; Sequence 109, Application US/10085959
; Publication No. US20030165870A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; APPLICANT: Burland, Valerie D.
; TITLE OF INVENTION: No. US20030165870A1e1 Sequence of E. Coli CFT073
; FILE REFERENCE: 960296.97648
; CURRENT APPLICATION NUMBER: US/10/085.959
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/242,412
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 109
; LENGTH: 76804
; TYPE: DNA
; ORGANISM: Escherichia coli

FEATURE:
; NAME/KEY: misc feature
; LOCATION: (39751)..(39751)
; OTHER INFORMATION: Unsure
; NAME/KEY: misc feature
; LOCATION: (49372)..(49372)
; OTHER INFORMATION: Unsure
; NAME/KEY: misc feature
; LOCATION: (66539)..(66539)
; OTHER INFORMATION: Unsure
; NAME/KEY: misc feature
; LOCATION: (73693)..(73693)
; OTHER INFORMATION: Unsure
; NAME/KEY: misc feature
; LOCATION: (73705)..(73705)
; OTHER INFORMATION: Unsure
; NAME/KEY: misc feature
; LOCATION: (73737)..(73737)
; OTHER INFORMATION: Unsure
; NAME/KEY: misc feature
; LOCATION: (73759)..(73759)
; OTHER INFORMATION: Unsure
; NAME/KEY: misc feature
; LOCATION: (73761)..(73761)
; OTHER INFORMATION: Unsure
; NAME/KEY: misc feature
; LOCATION: (73774)..(73774)
; OTHER INFORMATION: Unsure
; NAME/KEY: misc feature
; LOCATION: (73781)..(73781)
; OTHER INFORMATION: Unsure
; NAME/KEY: misc feature
; LOCATION: (73798)..(73798)
; OTHER INFORMATION: Unsure
; NAME/KEY: misc feature
; LOCATION: (73802)..(73802)
; OTHER INFORMATION: Unsure

US-10-085-959-109

Query Match 99.5%; Score 2079.8; DB 15; Length 76804;
Best Local Similarity 99.7%; Pred No. 0;
Matches 2084; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy 1 ATGCGAATAACCACTCTCGCTTCCTAGTCAATTCCTCTCGGATTTTCAGCCAGCAGC 60
Db 44573 ATGCGAATAACCACTCTCGCTTCCTAGTCAATTCCTCTCGGATTTTCAGCCAGCAGC 44514
Qy 61 ATAGCTGCTGCAGAGGATGATGATGCTCTCGGATCCGGCTATGAGAAAAGCTGACT 120
Db 44513 ATAGCTGCTGCAGAGGATGATGATGCTCTCGGATCCGGCTATGAGAAAAGCTGACT 44454
Qy 121 AACGACGCGCCAGTGTCTCTGATTAGCCAGGAGGAATTGACGTCCAGCCAGTACCAC 180
Db 44453 AACGACGCGCCAGTGTCTCTGATTAGCCAGGAGGAATTGACGTCCAGCCAGTACCAC 44394
Qy 181 GATCTGGCGGAGGCTCTGAGATCAGTAGAGGTGTGGATGTTGAAAGTGGTACGGGTAAA 240
Db 44393 GATCTGGCGGAGGCTCTGAGATCAGTAGAGGTGTGGATGTTGAAAGTGGTACGGGTAAA 44334
Qy 241 ACCGAGGCGCTCGAAATCAGCATCCGAGGAATGCCAGCCAGTTACACGCTGATCTGATT 300
Db 44333 ACCGAGGCGCTCGAAATCAGCATCCGAGGAATGCCAGCCAGTTACACGCTGATCTGATT 44274
Qy 301 GATGCTGCTGCAGAGGAGGAGGAGTACGTGACTCCCAACGGTTTCTCCCATGAT 360
Db 44273 GATGCTGCTGCAGAGGAGGAGGAGTACGTGACTCCCAACGGTTTCTCCCATGAT 44214
Qy 361 ACCGAGGCGCTCGAAATCAGCATCCGAGGAATGCCAGCCAGTTACACGCTGATCTGATT 420
Db 44213 ACCGAGGCGCTCGAAATCAGCATCCGAGGAATGCCAGCCAGTTACACGCTGATCTGATT 44154
Qy 421 TCACACTGTATGGCTCTGATCGATGGCGGTGGTGAATATCATATACCAAGAAAT 480
Db 44153 TCACACTGTATGGCTCTGATCGATGGCGGTGGTGAATATCATATACCAAGAAAT 44094
Qy 481 GCAGACAAATGCTCTCTTCCTGCAATGCAGAGGCTGAATCTCAGGAAAGCAACAAATG 540
Db 44093 GCAGACAAATGCTCTCTTCCTGCAATGCAGAGGCTGAATCTCAGGAAAGCAACAAATG 44034
Qy 541 GGTACAGCAGCAGGCTTAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTCTGTGAC 600
Db 44033 GGTACAGCAGCAGGCTTAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTCTGTGAC 43974
Qy 601 CTGACGCTACGGGTAGACACAAAGCCTGAGGTTATCGGTCACTACATCAGTACGAT 660
Db 43973 CTGACGCTACGGGTAGACACAAAGCCTGAGGTTATCGGTCACTACATCAGTACGAT 43914
Qy 661 ACAGCAGGCGCAGTATTCCTTATCCAGGAGTCAAGAAATTAATCTTGTGCACTG 720
Db 43913 ACAGCAGGCGCAGTATTCCTTATCCAGGAGTCAAGAAATTAATCTTGTGCACTG 43854
Qy 721 CTTGACTGGAAGCGCTCGGAGCAGGATGCTCTGTTTGTATGATACCAACCCGCGAG 780
Db 43853 CTTGACTGGAAGCGCTCGGAGCAGGATGCTCTGTTTGTATGATACCAACCCGCGAG 43794
Qy 781 CGTTATGATAACCGGATGGCAACTGGGAGTCTGACGGGGGATATCACCGGACCTG 840
Db 43793 CGTTATGATAACCGGATGGCAACTGGGAGTCTGACGGGGGATATCACCGGACCTG 43734
Qy 841 CGCTATGAGCGAAACAAATTTTCAGCTGGCTATGATCATACTTTCACCTTCGGAACATGG 900
Db 43733 CGCTATGAGCGAAACAAATTTTCAGCTGGCTATGATCATACTTTCACCTTCGGAACATGG 43674
Qy 901 AATCGTATCTGAATCTGGAACGAGACAGAAATTAAGGTCGTGAGTTTACCGCTGTA 960
Db 43673 AATCGTATCTGAATCTGGAACGAGACAGAAATTAAGGTCGTGAGTTTACCGCTGTA 43614
Qy 961 CTGAAGCGCAGCAAAATGGGGGCTTGGCGGTTCAGCCCGGAGCTTAAGGAATCGAACCTT 1020
Db 43613 CTGAAGCGCAGCAAAATGGGGGCTTGGCGGTTCAGCCCGGAGCTTAAGGAATCGAACCTT 43554


```
QY 1021 ATCTGTAATTCATTACTGCTTACCCCTCTGGGAGAAATCTCATCTGGTTACGGTGGGGGCG 1080
Db 43553 ATCTGTAATTCATTACTGCTTACCCCTCTGGGAGAAATCTCATCTGGTTACGGTGGGGGCG 43494
QY 1081 GAGTTTCAGAGCTCGTCCATGAAGACGGAGTTGCTCTTCCAGCAGACAGGTGAACACTTTC 1140
Db 43493 GAGTTTCAGAGCTCGTCCATGAAGACGGAGTTGCTCTTCCAGCAGACAGGTGAACACTTTC 43434
QY 1141 CGCAGAAAGCTGCTGCTGATTTGCTGAGGATGAGTGCATCTCACGGATCACTTCG 1200
Db 43433 CGCAGAAAGCTGCTGCTGATTTGCTGAGGATGAGTGCATCTCACGGATCACTTCG 43374
QY 1201 CTGACTGCGGCGAGCGCTGATGAACATCATGAGCAATTCGGGGACACTTCACTCCGGCT 1260
Db 43373 CTGACTGCGGCGAGCGCTGATGAACATCATGAGCAATTCGGGGACACTTCACTCCGGCT 43314
QY 1261 GCATATCTGGTCTGGGATGTGGCAGATGCTGGACGCTGAAGAGCGGTGTGACACGGGA 1320
Db 43313 GCATATCTGGTCTGGGATGTGGCAGATGCTGGACGCTGAAGAGCGGTGTGACACGGGA 43254
QY 1321 TATAAGGCAACCAAGATGGGCGAGCTTACATAAAGGATTTAGTGGTGTGTCGGGCGAGGA 1380
Db 43253 TATAAGGCAACCAAGATGGGCGAGCTTACATAAAGGATTTAGTGGTGTGTCGGGCGAGGA 43194
QY 1381 AAAACAAATCTACTTGGTAAACCCGACCTGAAGCCGGAAGAGCGTCAAGTTATGAGCT 1440
Db 43193 AAAACAAATCTACTTGGTAAACCCGACCTGAAGCCGGAAGAGCGTCAAGTTATGAGCT 43134
QY 1441 GGGGTCTATTAGATAACCCGCGCTCTGAATGCCAATGTACAGAGTTTATGACTGAC 1500
Db 43133 GGGGTCTATTAGATAACCCGCGCTCTGAATGCCAATGTACAGAGTTTATGACTGAC 43074
QY 1501 TTCTCCAAACAAGATTGTCTCTTATTCATAAATGATAACCAATAGCTATGTAACAGC 1560
Db 43073 TTCTCCAAACAAGATTGTCTCTTATTCATAAATGATAACCAATAGCTATGTAACAGC 43014
QY 1561 GGAAGGCGCGGTGACAGGTGTGAATTTGCGGACACTTGCCTGTGTGTCAGAGAT 1620
Db 43013 GGAAGGCGCGGTGACAGGTGTGAATTTGCGGACACTTGCCTGTGTGTCAGAGAT 42954
QY 1621 GTCAAGCTGTCACTGAATPACACCTGGACCCGAACTGAACCAACGATGGTGAATAACAA 1680
Db 42953 GTCAAGCTGTCACTGAATPACACCTGGACCCGAACTGAACCAACGATGGTGAATAACAA 42894
QY 1681 GGTGCGCGCTGAGTTATACCCCTGAAACACATGTTGAATGCAAACTGAACTGCGCAGATC 1740
Db 42893 GGTGCGCGCTGAGTTATACCCCTGAAACACATGTTGAATGCAAACTGAACTGCGCAGATC 42834
QY 1741 ACCGAAGAGTGGCATCATGGTGGTGGTCCGCTTATCGGGGAAACACCACTTTTACC 1800
Db 42833 ACCGAAGAGTGGCATCATGGTGGTGGTCCGCTTATCGGGGAAACACCACTTTTACC 42774
QY 1801 CAGAAATATTCCTCACTGAGCGCTGTACAGAAAGTGTATGATGAGAAAGGAGAAATAC 1860
Db 42773 CAGAAATATTCCTCACTGAGCGCTGTACAGAAAGTGTATGATGAGAAAGGAGAAATAC 42714
QY 1861 CTGAAGCTGACCGGTGGATGCGAGTGTGCTGCGGAAGATGACGGATGCCCTGACG 1920
Db 42713 CTGAAGCTGACCGGTGGATGCGAGTGTGCTGCGGAAGATGACGGATGCCCTGACG 42654
QY 1921 CTGAATGTGCGGTGAATAACCTGTCTCAACAGGATTAACAGTACGCTGAGCTGTACAGT 1980
Db 42653 CTGAATGTGCGGTGAATAACCTGTCTCAACAGGATTAACAGTACGCTGAGCTGTACAGT 42594
QY 1981 GCGGTAGAGTACCTGTATGCGGTGATTAATCTCCAGACGGATCATCAACACAGGA 2040
Db 42593 GCGGTAGAGTACCTGTATGCGGTGATTAATCTCCAGACGGATCATCAACACAGGA 42534
QY 2041 TATGTGATACCTGACGAAATTAATCTGAGTGTGCTGGAATCACTCACTGATCTGA 2091
Db 42533 TATGTGATACCTGACGAAATTAATCTGAGTGTGCTGGAATCACTCACTGATCTGA 42483
```

```
RESULT 3
US-10-238-075-1076
; Sequence 1076, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolates
; of E. coli, and biological uses of these polynucleotides and of their
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; PRIOR FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1076
; LENGTH: 2046
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-238-075-1076
```

Query Match 6.0%; Score 125.6; DB 15; Length 2046;

Best Local Similarity 46.3%; Pred.No. 4.1e-31; Mismatches 759; Indels 33; Gaps 7;

Matches 684; Conservative 0;

```
QY 72 AGAGGATGTGATGATTGTCTCGGCATCCGGCTATGAGAAAAAGCTGACTAAGCGACCCG 131
Db 87 AGAAGATACGCTTGTGTTACTGCTCCGGTTCACTCAGCAGCTCAGAAATGCCCGGC 146
QY 132 CAGTGTTCGTGATTAGCCAGGAGAAATGCAGTCCAGCCAGTACCAGATCTGGCGA 191
Db 147 CAGTGTCTCAGTCATTACTTCAGAACCACTGCAAAAAACCGGTTTCAGATCTGGTCA 206
QY 192 GGCTCTGAGATCAGTAGAGGGTGTGGATGTGAAAGTGTACGGGTAAACCCGAGGGCT 251
Db 207 TG-----CAGTAAAGATGTTGAAGGATAGTATCACTGGTGGGAATGAAAAAC 257
QY 252 GAAATCAGATCCGAGGAATGCCAGCAGTTACCGTGTACTGATGATGATGATGTTGCG 311
Db 258 GGATATCAGTATACGTGCTAAGTGGCGATTAACCGTGTATTCGTCATGGCAGCG 317
QY 312 TCAGGCGGAAGCAGTGACGTGACTCCCAACGGTTTTCTGCCATGAATACCGGGTTCA 371
Db 318 TCAGAGCG---GTCGGGAATCCAGACCAACCGCAGCGCGGTTTGAACCGGATTA 374
QY 372 GCCCTCTGCGCGCAATTAAGCGTATTGAGGTTATCAGGGGGCCGATGTCACACTGTA 431
Db 375 CCCTCCTGTGGAAGCAATTTGAACCGCATTCGATGATCCGTCCTATGCTTCCCTGTA 434
QY 432 TGGCTCTGATGCGTGGCGGTGTGGTGAATATCAATACCAAGAAAGATGCAGACAAATG 491
Db 435 TGTTCTGATGCGATCCGAGGGGTCAATTAATATCAACCAACCCAGTTAATACCAAC 494
QY 492 GCTCTCTCCGTCATATGCA---GGGCTGAATCTGCAGGAAGCAACAAATGGGGTAACAG 548
Db 495 ATGGGATGCGGTACTTTGGACTTTGGGGGATTAATCAGGAACATGGGAAATTTGGTAATC 554
QY 549 CAGCAGATTAAATTTCTGAGCAGTGGTCCCTTGTGGATGATTTCTGTCAGCTCAGGT 608
Db 555 AACCACAAATGATCTTCTATCTGTGAGGCCCATTTGATTAAGGATAAACTTGGTCTTCAGCT 614
QY 609 ACSCGCTAGCACACACACGCTCAGGGTTTCATCGGTCACTCACTGAGCGATACAGCAGG 668
Db 615 ATATGGAGGAATGAACATACCGAAGS-----AAGATAGTATCTCTCAGGA-ACACCGGC 668
QY 669 CAGCGTATTCCTTATCCACCGAGTACAGAAATTAATCTTGGTGAACGCTCTTACTG 728
Db 669 AAAAGATAATAAGATATAACCGCAACGCTCCAGTTTATCCGACTGAAGACAGAGTT 728
QY 729 GAAGGCGTCGAGCAGGATGTCTCTGTTTTCATATGGATACACCCCGCAGCGTTATGA 788
Db 729 TGTGTTTGAATATGGAATAAATAACAGGTGATATACACCTGGTGTCTCTCGA 788
```



```
Qy 789 TAACCGGATGGGCAACTGGGAGTCTGACGGGGGAGATGACCGGACCCCTGGCTATGA 848
Db 789 TGCCTGGACTATGGGGGAAATCTTAAACAAACCAACAGTAAAGAGAAACGCAATATTC 848
Qy 849 GCGAAACAAATTTTCACTGCTATGATCATCTTTTCACTTTCGGAACATGGAAATCGTA 908
Db 849 ACATGATGCTATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 902
Qy 909 TCTGAATCGGAACGAGACAGAAATAAAGTTCGTGAGCTTGTACGCGAGTCTATGAAAGC 968
Db 909 AATTGCTGTTTATCAGGAGAAAGTTATTCGTGAGCTTAAATCAGGTAAAGAAAGATAATA 962
Qy 969 CGACAAATGGGGCTTCCCGGTGAGCGCGGAGCTTAAGGAATCGAACCTTATCTCTGAA 1028
Db 969 TAATCATTTGGGATCTTAAATACGAGTCAAGAAACCGGAATTAACCAACATATAGTA 1022
Qy 1029 TTTAATTACTGTTACCCCTCTGGGAGAAATCTCATCTGTTTACGTTGGGGGCGAGTTTCA 1088
Db 1029 TGCATAAGTACGCAATTTCTGCCGAAATGTACTGACCATCGAGGTCAAATTCAGCA 1082
Qy 1089 GAGCTCGTCCATGAAGACGAGTGTCTTGCAGACACAGGTGAAA-----CTTCCG 1142
Db 1089 TGCAGAGTCCGATGACTCAGCCAGCGGTAAAGAAACGACAGAAACACAGTCTGTTTC 1142
Qy 1143 GCAGAAAGCTGTCGCTGATTTTGTGAGGATGAGTGGCATCTCACGGATGCACTTGGCT 1202
Db 1143 AATTAAACAGAACTGTTTATAGAAATGAATATGACGACAGATCTCTCGCCCT 1202
Qy 1203 GACTCGCGGACGCTATGAACATCATGACAAATCGGGGACACTTCAGTCGCGCTGC 1262
Db 1203 GACTGGAGGACTGGTCTCGAATATCATGAATCTATGCGAGTTACTGGAATCCAAGATT 1262
Qy 1263 ATATCTGCTGGGATGTCGAGATGCTGAGCTGAAAGCGGTGACCGGGATA 1322
Db 1263 GTAGCTGTTTAACTGACCGAATATCTCACACTCAAGGGGGGATCGCAAGCATT 1322
Qy 1323 TAAGGCAACCAAGATGGGCGAGTACATAAAGGAATAGTGGTGTCCGGGACGGAAA 1382
Db 1323 TCGGCTCTTCAATCTGAGGTGAGTCTCGGATTTGGAACACTGACGAGGTGGTG 1382
Qy 1383 AACAAATCTACTTGTAAACCCGACCTGAACCCGGAAGAGCGTCACTATGAGGCTGG 1442
Db 1383 CTCTATTATGATGGAACAGGGACCTGAAACCGGAGACAGTGTAAACCGAGAGATCGG 1442
Qy 1443 GGTGATTACGATAACCCCGCGTCTGAATGCCAATGTCAAGGTTTATGACTGACTT 1502
Db 1443 TATTATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1502
Qy 1503 CTCACCAAGATGCTCTTATTCATTAATGATAA 1538
Db 1503 TAAAAATAAGTTGACCACTTACGATATAGGTACAAA 1538

RESULT 4
US-10-238-075-1049
; Sequence 1049, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of their
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-03-10
; PRIORITY APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1049
; LENGTH: 12943
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-238-075-1049
```

```
Query Match 6.0%; Score 125.6; DB 15; Length 12943;
Best Local Similarity 46.3%; Pred. No. 1.3e-30;
Matches 684; Conservative 0; Mismatches 759; Indels 33; Gaps 7;

Qy 72 ACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 131
Db 10799 AGAAGATACGCTTGTGTTTACTGCTCCGGTTCCTACAGCAGCTCAGAAATGCCCCGC 10858
Qy 132 CAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 191
Db 10859 CAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10918
Qy 192 GCCTCTGAGATCAGTAGAGGGTGTGATGTTTGAAGTGTGATGATGATGATGATGATGAT 251
Db 10919 TG-----CAGTAAAGATGTTGAAGGATTAAGTATCACTGCTGGGATGAAAC 10969
Qy 252 GGAATACAGATCCAGGGAATCCAGCCAGTTCACGCTGATGATGATGATGATGATGATGAT 311
Db 10970 GGATATCAGTATACGTTGTTAAGTGGGATTAACAGCTGATTCCTGGTATGAGACGAG 11029
Qy 312 TCAGGCGGAAGCAGTACGCTGACTCCCAACGGTTCCTGCTCATGATACCGGTTTCAT 371
Db 11030 TCAGAGCG---GTCCGGAATCCAGAACGCGAGCGCGGTTTGAAGCCGGATTAT 11086
Qy 372 GCGCCCTCTGGCGGCAATGAGCTGATGAGGTATATCAGGGGCGGATGTCACACTGTA 431
Db 11087 CCTCTCTGTGGAAGCAATTTGAACGCAATTTGAAGTATCGGTGCGCTATGCTTCTCCTGTA 11146
Qy 432 TGGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 491
Db 11147 TGGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 11206
Qy 492 GCTCTCTCCGTCATGCA---GGGCTGAATCTGAGGAAGCAACAATGGGTAACAG 548
Db 11207 ATGGGATGCTGATCTTGGACTTGGGGGATTAATCAGGAACATGGAATTTGGTAAC 11266
Qy 549 CAGCCAGTTAATTTCTGAGCAGTGTCTCCCTTGTGATGATGATGATGATGATGATGATGAT 608
Db 11267 AACCAAAATGACTTCTATCTGTCAGGCGCAATGATTAAGGATAAACTTGGTCTTCAGCT 11326
Qy 609 ACGCGGTACACACACAGCGGTGAGGTTTCATCGGTGATGATGATGATGATGATGATGATGAT 668
Db 11327 ATATGAGGAATGAACTATCGCAAGG-----AAGATAGTATCTCTCAGGGA-ACACCGGC 11380
Qy 669 CAGCGCTATTTCTTATCCCAAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 728
Db 11381 ABAAGATAATAAGAAATATAACGCAACGCTCCAGTTTACTCCGACTGAAAGCCAGAAAT 11440
Qy 729 GAAGCGCTCGAGCAGGATGCTCTGTTTGTATGATGATGATGATGATGATGATGATGATGAT 788
Db 11441 TGTTTTGAATATGGAATAAATAACAGGTGATGATGATGATGATGATGATGATGATGATGAT 11500
Qy 789 TAACCGGATGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCCCTGGCTATGA 848
Db 11501 TGCCTGCACTATGCGGGGAAATCTTAAACAAACCAACAGTAAAGAGAAACGCAATATTC 11560
Qy 849 GCGAAACAAATTTTCACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 908
Db 11561 ACCTAGTCACTGGGTAGCAGCAT-----GGAATGCCAGGCGAAATATCTGCATCTCGA 11614
Qy 909 TCTGAACTGGAACGACACAGAAATAAAGTCTGTGAGCTTGTACGCACTGATGATGATGATGAT 968
Db 11615 AATTGCTGTTTATCAGGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 11674
Qy 969 CGACAAATGGGGCTTCCCGGTGAGCGCGGAGTGTAAAGGAATCGAACTTATCTCTGAA 1028
Db 11675 TAATCATTTGGATCTTAAATTAAGTCAAGAAACCGGAATAACCAACATATATAGTA 11734
Qy 1029 TTTAATTACTGCTTACCCCTCTGGGAGAAATCTCATCTGTTTACGTTGGGGGCGAGTTTCA 1088
Db 11735 TGCAAAAGTACGCGCAATTTCTGCCGAAATATGATGATGATGATGATGATGATGATGATGAT 11794
```


QY 1089 GAGCTCGTCCATGAAAGCGGAGTGTCTCTGCCAGCACAGGTGAAA-----CTTTCCG 1142
Db 11795 TGCAGAGCTCCGATGATCTGACCCACGGGTAAAAAACACAGAAACACAGCTCTGTTTC 11854
QY 1143 GCAGAAAGCTGGTGGTATTTGCTGAGGATGAGTGGCATCTCACGGATGCACTTGCCT 1202
Db 11855 AATTAAACAGAAAGCTGTTTTATAGAAATGAATATGACGCAACGGATCTCTCGCCT 11914
QY 1203 GACTCGGCGAGCGCTATGACATCATGACAAATTCGGGACACATTCAGTCCGCTGC 1262
Db 11915 GACTGAGGACTCGCTGCTGAAATCATGAATCTATGCGAGTTACTGGAATCCAGATT 11974
QY 1263 ATATCTGCTGCTGGATGTGGCAGATGCTGAGCGCTGAAAGCGGTGTGACACCGGATA 1322
Db 11975 GTACGCTGTTTATACTGACCGATAATCTCACACTCAAGGGGGATCGCAAGACATT 12034
QY 1323 TAAGGACCCAGAAATGGGCGAGTACATTAAGGATTTAGTGTGTCTGGGCGAGGAAA 1382
Db 12035 TCGGGCTCTTCAATTCGTGAGGTGAGTCTCTGGATTGGAACTGACGCGAGGTGTGC 12094
QY 1383 AACAAATCTACTTGGTAACCCGACCTGAAGCCGGAAGAGACGCTCAGTTATGAGCTGG 1442
Db 12095 CTCATTATGATGAAACAGGACCTGAAACCGGAGACCATGTAAACGAGATCGG 12154
QY 1443 GGTGATTAAGATAACCCCGGCTGTAATGCCAATGTACAGGTTTTTATGACTGATT 1502
Db 12155 TATTATTATAGTAATGATAGTGTGTTTTTCGGCGAGCGACGCTGTTTAACTGATT 12214
QY 1503 CTCACACAGATTGCTCTTATTCATTAATGATAA 1538
Db 12215 TAAAAATAAGTTGACCAAGTTACGATATAGGTACAAA 12250

RESULT 5

US-09-764-875-112
; Sequence 112, Application US/09764875
; Publication No. US2004001896A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ202
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 112
; LENGTH: 758
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-875-112

Query Match 3.1%; Score 64.2; DB 11; Length 758;
Best Local Similarity 57.1%; Pred. No. 2.4e-10;
Matches 117; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 388 ATTGACGCTATTGAGTTATCAGGGGCGGATGTCACACTGTATGGCTCTGATCGCATG 447
Db 427 ATTGAACGCTATTGAGTTCTGCGTGTCCGGCAGCTGCGGTTATGGCAACGGCGCGC 486
QY 448 GCGGTGTGGTGAATATCATTTACCAAGAAAGATGCAGACAAATGGCTCTCTTCGGTCAAT 507
Db 487 GCGCGGTGGTTAATCATTTACCAAAAAAAGCAGCGCGAGTGGCAGGCTCTCTGGAC 546
QY 508 GCAGGCTGAATCTCAGGAAGCAACAAATGGGTTACAGCAGCAGTTAAATTTCTGG 567
Db 547 GCATATTTCAATGCGCCAGAACATAAAGAGGAAGTGCACCAACGCACTAACTTTAGC 606
QY 568 AGCAGTGTCCCTTGTGGATGATT 592
Db 607 CTGACCGTCCGCTGGCGACGAAT 631

RESULT 6

US-09-764-868-94/c
; Sequence 94, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT332
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 94
; LENGTH: 822
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (795)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (804)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-94

Query Match 3.1%; Score 64.2; DB 9; Length 822;
Best Local Similarity 57.1%; Pred. No. 2.6e-10;
Matches 117; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 388 ATTGACGCTATTGAGTTATCAGGGGCGGATGTCACACTGTATGGCTCTGATCGCATG 447
Db 324 ATTGAACGCTATTGAGTTCTGCGTGTCCGGCAGCTGCGGTTATGGCAACGGCGCGC 265
QY 448 GCGGTGTGGTGAATATCATTTACCAAGAAAGATGCAGACAAATGGCTCTCTTCGGTCAAT 507
Db 264 GCGGCGGTGGTTAATCATTTACCAAAAAAAGCAGCGGCGAGTGGCAGGCTCTCTGGAC 205
QY 508 GCAGGCTGAATCTCAGGAAGCAACAAATGGGTTACAGCAGCAGTTAAATTTCTGG 567
Db 204 GCATATTTCAATGCGCCAGAACATAAAGAGGAAGTGCACCAACGCACTAACTTTAGC 145
QY 568 AGCAGTGTCCCTTGTGGATGATT 592
Db 144 CTGACCGTCCGCTGGCGACGAAT 120

RESULT 7

US-09-764-868-510
; Sequence 510, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT332
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 510
; LENGTH: 686
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (19)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (28)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-510


```

Query Match      2.8%; Score 57.6; DB 9; Length 686;
Best Local Similarity 58.0%; Pred. No. 4.2e-08;
Matches 102; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 388 ATTGAGCGTATTGAGTTATTCAGGGGCGGATGCCACACTGATGGCTCTGTATGGCATG 447
Db 499 ATTGAACGTTAGTTCTGCGTGGTCCGGCAGCTGGCGTTATGCAACGGCGGGCG 558

QY 448 GCGGGTGTGGTATATCATTCAGAGAAAGATGCAGACAAATGGCTCTCTTCGTCAT 507
Db 559 GCGGCGGTGGTTAATCATATTACCAAAAAAGGCGAGTGGCAGCGCTCTCTGGGAC 618

QY 508 GCAGGCGTGAATCGCAGGAAGCAACAATGGGTTAACAGCAGCCAGTTTAATTT 563
Db 619 GCATATTTCATGCGCCAGAACATAAAGAGGAAGTGCACCAACGCATTAATT 674

RESULT 8
US-09-764-875-424
; Sequence 424, Application US/09764875
; Publication No. US20040018969A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P0202
; CURRENT APPLICATION NUMBER: US/09/764,875
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1249
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 424
; LENGTH: 686
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (19)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (28)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-875-424

Query Match      2.8%; Score 57.6; DB 11; Length 686;
Best Local Similarity 58.0%; Pred. No. 4.2e-08;
Matches 102; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 388 ATTGAGCGTATTGAGTTATTCAGGGGCGGATGCCACACTGATGGCTCTGTATGGCATG 447
Db 499 ATTGAACGTTAGTTCTGCGTGGTCCGGCAGCTGGCGTTATGCAACGGCGGGCG 558

QY 448 GCGGGTGTGGTCAATATCATTCAGAGAAAGATGCAGACAAATGGCTCTCTTCGTCAT 507
Db 559 GCGGCGGTGGTTAATCATATTACCAAAAAAGGCGAGTGGCAGCGCTCTCTGGGAC 618

QY 508 GCAGGCGTGAATCTCGAGGAAGCAACAATGGGTTAACAGCAGCCAGTTTAATTT 563
Db 619 GCATATTTCATGCGCCAGAACATAAAGAGGAAGTGCACCAACGCATTAATT 674

RESULT 9
US-09-864-408A-851
; Sequence 851, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shinkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24

```

```

; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 851
; LENGTH: 296
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-864-408A-851

Query Match      2.5%; Score 51.8; DB 11; Length 296;
Best Local Similarity 60.1%; Pred. No. 2.4e-06;
Matches 86; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 1372 GGGCAGGGGAAAAAATACTTGTGTAACCCCGACCTGAGCCGGAAGAGAGAGCTCAGT 1431
Db 7 GGTGGGCGACCTCAATCATTTATCGGCAACCCGACCTGCGAGCCAGAGAGAGCTGAGC 66

QY 1432 TATGAGGCTGGGGTGTATTACGATAACCCCGCGGTCTGTAATGCCAAATGTACAGTTT 1491
Db 67 GAAGAGATCGGCGTGTATGTGGGATAACCAACAGGCGCTGAACCGCGGTAAACCTGTTC 126

QY 1492 ATGACTGACTTCTCCAAACAAGAT 1514
Db 127 AACACCGACTTCAAGACAAGAT 149

RESULT 10
US-10-297-465A-1
; Sequence 1, Application US/10297465A
; Publication No. US20040142413A1
; GENERAL INFORMATION:
; APPLICANT: Simpson, Andrew
; APPLICANT: Reinach, Fernando
; APPLICANT: Setubal, Joao
; APPLICANT: Mediana, Joao
; APPLICANT: Arruda, Paulo
; TITLE OF INVENTION: Isolated Genome of Xylella fastidiosa and Uses Thereof
; FILE REFERENCE: P0202 US (10213376)
; CURRENT APPLICATION NUMBER: US/10/297,465A
; CURRENT FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: PCT/IB01/01618
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,906
; PRIOR FILING DATE: 2001-06-17
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465A-1

Query Match      2.5%; Score 51.4; DB 17; Length 2731748;
Best Local Similarity 54.0%; Pred. No. 0.00096;
Matches 128; Conservative 0; Mismatches 106; Indels 3; Gaps 1;

QY 240 AACCGAGGGCTGGAATCAGCATCCGAGGAATCCAGCAGTACACGCTGATCTGAT 299
Db 347778 AATGGGTGGAGTGTATCTCGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 347837
QY 300 TGATGGTGTTCGTGAGGGGGAAGCAGTACGTGACTCCCAACGGTTTTTCTCCCATGAA 359
Db 347838 TAACGG--TCGTGATGCTCTTCTTACGGTTGGCCGACGATGTCAGAAAGTTGTCG 347894

QY 360 TACCGGGTTCATGCCCGCTCTGCGCCCATTCAGCGTATGAGTTATCAGGGGCGCAT 419
Db 347895 TGATTTAGTAGCATCCGATGCGGTGAGCGTGTGAAGTTCTGAAGAGCGGTGC 347954
QY 420 GTCCACACTGTATGGTCTGTGATGCGATGGCGGTGTGGTGAATATCATTTACCAGAAA 476
Db 347955 GTCTCAATTTAGGATCTGACCGGATTCCTGGTGTAGTCAATATCATTTTGGGTAA 348011

RESULT 11

```



```
US-10-238-075-526
; Sequence 526, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 526
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-238-075-526

Query Match      2.3%; Score 48.6; DB 15; Length 1962;
Best Local Similarity 54.9%; Pred. No. 9.8e-05;
Matches 96; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 378 TCTGGCGCCGATTGAGCGTATTGAGTTATCAGGGGGCCGATGTCCACACTGTATGGCTC 437
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 363 TCGGCACTCGTCAACGATATCGAAGTTGTCCGGGGCCCAAACTCCAGTCTGTACGGCAG 422
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 438 TGATGCGATGGCGGTGTGGTGAATATCATTACCAAGAAAGATGCAGACAAATGGCTCTC 497
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 423 TGGCGGCTGGAGGTGTAGTGCATTCAGAACTCCGATGCAGCAGATTTCTTCCCCC 482
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 498 TTCCGTCAATGAGCGCTCAATCTCAGGAAAGCAAAATGGGTAAACAGCAGC 552
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 483 CGGAGAGACAAACGGTTTAAGTCTGTGGGAAATATCGCCAGTGTGACCAAGC 537
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
US-10-238-075-507/c
; Sequence 507, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 507
; LENGTH: 1264
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-238-075-507

Query Match      2.3%; Score 48.6; DB 15; Length 12264;
Best Local Similarity 54.9%; Pred. No. 0.0003;
Matches 96; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 378 TCTGGCGCCGATTGAGCGTATTGAGTTATCAGGGGGCCGATGTCCACACTGTATGGCTC 437
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 9831 TCGGCACTCGTCAACGATATCGAAGTTGTCCGGGGCCCAAACTCCAGTCTGTACGGCAG 9772
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 438 TGATGCGATGGCGGTGTGGTGAATATCATTACCAAGAAAGATGCAGACAAATGGCTCTC 497
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 9771 TGGCGGCTGGAGGTGTAGTGCATTCAGAACTCCGATGCAGCAGATTTCTTCCCCC 9712
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 498 TTCCGTCAATGAGCGCTCAATCTCAGGAAAGCAAAATGGGTAAACAGCAGC 552
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 9711 CGGAGAGACAAACGGTTTAAGTCTGTGGGAAATATCCGCAAGTGTGACCAAGC 9657
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 13
US-10-085-959-11
; Sequence 11, Application US/10085959
; Publication No. US20030165870A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; APPLICANT: Burland, Valerie D.
; TITLE OF INVENTION: No. US20030165870A1 Sequence of E. Coli CFT073
; FILE REFERENCE: 960296.97648
; CURRENT APPLICATION NUMBER: US/10/085,959
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/242,412
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 23654
; TYPE: DNA
; ORGANISM: Escherichia coli
; NAME/KEY: misc feature
; LOCATION: (12189)..(12189)
; OTHER INFORMATION:
; NAME/KEY: misc feature
; LOCATION: (12189)..(12189)
; OTHER INFORMATION: Unsure
US-10-085-959-11

Query Match      2.3%; Score 48.6; DB 15; Length 23654;
Best Local Similarity 54.9%; Pred. No. 0.00046;
Matches 96; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 378 TCTGGCGCCGATTGAGCGTATTGAGTTATCAGGGGGCCGATGTCCACACTGTATGGCTC 437
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2433 TCGGCACTCGTCAACGATATCGAAGTTGTCCGGGGCCCAAACTCCAGTCTGTACGGCAG 2492
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 438 TGATGCGATGGCGGTGTGGTGAATATCATTACCAAGAAAGATGCAGACAAATGGCTCTC 497
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2493 TGGCGGCTGGAGGTGTAGTGCATTCAGAACTGCCGATGCAGCAGATTTCTTCCCCC 2552
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 498 TTCCGTCAATGAGCGCTCAATCTCAGGAAAGCAAAATGGGTAAACAGCAGC 552
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2553 CGGAGAGACAAACGGTTTAAGTCTGTGGGAAATATCCGCAAGTGTGACCAAGC 2607
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
US-10-194-163-462/c
; Sequence 462, Application US/10194163
; Publication No. US20020172976A1
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce Carter
; TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS POLYNUCLEOTIDES
; AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/194,163
; FILING DATE: 04-Nov-2002
; CLASSIFICATION: <Unknown>
```



```

; ATTORNEY/AGENT INFORMATION:
; NAME: Basu, Shantanu
; REGISTRATION NUMBER: 43,318
; REFERENCE/DOCKET NUMBER: 529282000101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5995
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 462
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2760 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: FORPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..2760
; SEQUENCE DESCRIPTION: SEQ ID NO: 462
US-10-194-163-462

Query Match      2.2%; Score 47; DB 13; Length 2760;
Best Local Similarity 57.4%; Pred. No. 0.00043;
Matches 105; Conservative 0; Mismatches 75; Indels 3; Gaps 1;

QY 388 ATTGAGCGTATTGAGTTATCAGGGGGCGGATGCCACACTGTATGGCTCTGATGGCATG 447
DB 1444 ATCGAACTGTGGAGGTAGTACGTGGTGAGGATCGGCGCTTGTACGGTCTCTTCTGCTATT 1385
QY 448 GCGGTGTGGTGAATATCATTTACCAGAAAGAAATGCAGACAAATGGCTCTCTTCGTCAT 507
DB 1384 CGCGAGTGGTGAATATCATTCACAGAACTTCTCACAATTTTTCACAT---TCAAT 1328
QY 508 CGAGGCTGAATCTGCAGAAAGCAACAAATGGGGTACAGCAGCCAGTTTAAATTTCTGG 567
DB 1327 GAATCTCTGAGCTTTACCGGTTTCAGCAAGCTGGATAACACAGCAACTTCAATGCCTCC 1268
QY 568 AGC 570
DB 1267 ATC 1265

RESULT 15
US-10-282-122A-12338
; Sequence 12338, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26

```

```

; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 12338
; LENGTH: 2061
; TYPE: DNA
; ORGANISM: Bacteroides fragilis
US-10-282-122A-12338

Query Match      2.1%; Score 43; DB 13; Length 2061;
Best Local Similarity 60.9%; Pred. No. 0.0084;
Matches 70; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 388 ATTGAGCGTATTGAGTTATCAGGGGGCGGATGCCACACTGTATGGCTCTGATGGCATG 447
DB 406 ATCGACCATATCGAATAGTCAAGAGAGCTTCTTGGCCCTCTATGTTCTCAAGCGATG 465
QY 448 GCGGTGTGGTGAATATCATTTACCAGAAAGAAATGCAGACAAATGGCTCTCTTCCG 502
DB 466 GCGGAGTTATCAACATCATTTACGCGTAAAGCCAAAAGAAATTCGAGGCTTCCG 520

Search completed: October 12, 2004, 18:30:26
Job time : 1007 secs

```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2004, 11:56:38 ; Search time 156 Seconds

(without alignments)
7438.476 Million cell updates/sec

Title: US-10-625-972-4

Perfect score: 2091
Sequence: 1 atcggaataaccactctggc.....cgctgaactatcagttctga 2091

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2091	100.0	2091	1	US-08-765-081-4
2	2091	100.0	2091	3	US-09-098-082-4
3	2091	100.0	2091	5	PCT-US95-06994-4
4	2091	100.0	8041	1	US-08-765-081-1
5	2091	100.0	8041	3	US-09-098-082-1
6	2091	100.0	8041	5	PCT-US95-06994-1
7	2089.4	99.9	87563	4	US-09-453-702B-57
8	127	6.1	2055	4	US-09-543-681A-736
9	105.4	5.0	2100	4	US-09-489-039A-962
10	105	5.0	1535	1	US-08-480-510-1
11	105	5.0	1535	5	PCT-US94-01780-1
12	96.8	4.6	2034	4	US-09-543-681A-3764
13	68.8	3.3	2280	4	US-09-489-039A-703
14	68.6	3.3	2495	4	US-09-658-113A-1
15	63.4	3.0	2388	4	US-09-489-039A-6223
16	61.8	3.0	1326	4	US-09-252-991A-15544
17	61.8	3.0	1437	4	US-09-252-991A-15295
18	61.2	2.9	439	1	US-08-480-510-10
19	61.2	2.9	439	5	PCT-US94-01780-10
20	61.2	2.9	517	1	US-08-480-510-9
21	61.2	2.9	517	5	PCT-US94-01780-9
22	61.2	2.9	534	1	US-08-480-510-8
23	61.2	2.9	534	5	PCT-US94-01780-8
24	56.4	2.7	2790	4	US-09-328-352-465
25	55.6	2.7	1116	4	US-09-252-991A-13319
26	55.6	2.7	2235	4	US-09-252-991A-12994
27	55.6	2.7	3486	4	US-09-252-991A-12477

28	54.4	2.6	2337	4	US-09-489-039A-1284	Sequence 1284, Ap
29	51.6	2.5	1845	4	US-09-307-973A-1	Sequence 1, Appli
30	50.2	2.4	933	4	US-09-252-991A-14912	Sequence 14912, A
31	50.2	2.4	939	4	US-09-252-991A-14458	Sequence 14458, A
32	48.4	2.3	1844	4	US-09-307-973A-2	Sequence 2, Appli
33	47.4	2.3	633	4	US-09-252-991A-8508	Sequence 8508, Ap
34	47.4	2.3	2712	4	US-09-252-991A-8406	Sequence 8406, Ap
35	47.4	2.3	3528	4	US-09-252-991A-8473	Sequence 8473, Ap
36	47	2.2	2760	4	US-09-221-017B-462	Sequence 462, App
37	46.8	2.2	1854	4	US-09-252-991A-14350	Sequence 14350, A
38	46.8	2.2	2343	4	US-09-252-991A-15342	Sequence 15342, A
39	45	2.2	300	4	US-09-489-039A-5016	Sequence 5016, Ap
40	43.8	2.1	1497	4	US-08-232-463-14	Sequence 14, Appl
41	43.4	2.1	7218	1	US-09-328-352-1816	Sequence 1816, Ap
42	42.8	2.0	2286	4	US-09-221-017B-916	Sequence 916, App
43	42	2.0	2182	4	US-09-252-991A-14718	Sequence 14718, A
44	41.8	2.0	405	4	US-09-252-991A-15032	Sequence 15032, A
45	41.8	2.0	1437	4	US-09-252-991A-15032	Sequence 15032, A

ALIGNMENTS

RESULT 1

US-08-765-081-4
; Sequence 4, Application US/08765081
; Patent No. 5798260
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible/Pentium
; OPERATING SYSTEM: MS-Windows 3.1
; SOFTWARE: Word for Windows-6.0

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/765,081
; FILING DATE: March 26, 1997

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995

APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994

ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997

REFERENCE/DOCKET NUMBER: CHOR-1-10286
; TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)

TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
; LENGTH: 2091 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
; DESCRIPTION: Corresponds to SEQ ID NO:1,
; DESCRIPTION: nucleotides 3036-5126
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

ORIGINAL SOURCE:
; ORGANISM: Escherichia coli 0157: H7
; STRAIN: 86-24 NAIR
; FEATURE:


```

Db      2041  TATGTGATACCTGACGCGAAATTAATCGGATGCTGCTGAACATATCAGTTCTGA 2091

RESULT 2
US-09-098-082-4
; Sequence 4, Application US/09098082
; Patent No. 6040421
;
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S.; Besser, T.E.; Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible/Pentium II
; OPERATING SYSTEM: MS-Windows 95
; SOFTWARE: Word for Windows-6.0
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,082
; FILING DATE:
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/765,081
; FILING DATE: March 26, 1997
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: CHOR-1-12402
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2091 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; DESCRIPTION: Corresponds to SEQ ID NO:1,
; DESCRIPTION: nucleotides 3036-5126
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli 0157: H7
; STRAIN: 86-24 NALR
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2088
;
US-09-098-082-4

Query Match 100.0%; Score 2091; DB 3; Length 2091;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2091; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ATCGGAATAACCACTCTGGGTTCCGTAGTCATTCCTGTCTCGGATTTTCAGCCAGCAGC 60
Db      1  ATCGGAATAACCACTCTGGGTTCCGTAGTCATTCCTGTCTCGGATTTTCAGCCAGCAGC 60
QY      61  ATAGCTGCTCAGAGGATGTGATGTTCTCGGCATCCGGTATGAGAAAAGCTGACT 120
Db      61  ATAGCTGCTCAGAGGATGTGATGTTCTCGGCATCCGGTATGAGAAAAGCTGACT 120
QY      121 AACGCGCCGCCAGTGTTCCTGTAATAGCAGGAGGAATTCAGTCCAGCCAGTACCAC 180

```


[illegible]

Qy	1501	TTCTCCAAAGAATGTGCTCTTATTCATAAATGATAACCAATAAGCTATGTAAACAGC	1560
Db	1501	TTCTCCAAAGAATGTGCTCTTATTCATAAATGATAACCAATAAGCTATGTAAACAGC	1560
Qy	1561	GGAAAGCCCGGTTGCACGGTGTGGAAATTGCGCGGCACATTGCGCGTGTGGTACAGAGAT	1620
Db	1561	GGAAAGCCCGGTTGCACGGTGTGGAAATTGCGCGGCACATTGCGCGTGTGGTACAGAGAT	1620
Qy	1621	GTCAACGCTGTCACTGAATTACACCTGCACCCGGAAGTGAAACAACGTCGATGTGATTAACAAA	1680
Db	1621	GTCAACGCTGTCACTGAATTACACCTGCACCCGGAAGTGAAACAACGTCGATGTGATTAACAAA	1680
Qy	1681	GGTGGCCGCTGAGTTATACCCCTGAAACACATGGTGAATGCGAAATCTGAACATCGGCAGATC	1740
Db	1681	GGTGGCCGCTGAGTTATACCCCTGAAACACATGGTGAATGCGAAATCTGAACATCGGCAGATC	1740
Qy	1741	ACCGAAGAGTGCATCATGCGTGGGTGCGCGTTATCGCGGGAAAAACACACGTTTCACC	1800
Db	1741	ACCGAAGAGTGCATCATGCGTGGGTGCGCGTTATCGCGGGAAAAACACACGTTTCACC	1800
Qy	1801	CAGAAATTATTCGTCACTGACCGCTGTACAGAAAGATGTATGATGAGAAAGAGAGATAC	1860
Db	1801	CAGAAATTATTCGTCACTGACCGCTGTACAGAAAGATGTATGATGAGAAAGAGAGATAC	1860
Qy	1861	CTGAAAGCCTGGACGGTGTGGATGACAGGTCTGTCTGTGGAAGATGACCGATGCCCTGACG	1920
Db	1861	CTGAAAGCCTGGACGGTGTGGATGACAGGTCTGTCTGTGGAAGATGACCGATGCCCTGACG	1920
Qy	1921	CTGAAATGCTCGGTGAATAACCTGTCTCAACAAGGATTACAGTGCAGTGCAGCCTGTACAGT	1980
Db	1921	CTGAAATGCTCGGTGAATAACCTGTCTCAACAAGGATTACAGTGCAGTGCAGCCTGTACAGT	1980
Qy	1981	GCCGGTAAGAGTACGCTGTATGCGCGGTGATTACTTCCAGACGGGATCATCAACAACAGGA	2040
Db	1981	GCCGGTAAGAGTACGCTGTATGCGCGGTGATTACTTCCAGACGGGATCATCAACAACAGGA	2040
Qy	2041	TATGTGATACCTGAGCGAAATTACTGCGATGTCGCTGAACCTATCAGTTCTCGA	2091
Db	2041	TATGTGATACCTGAGCGAAATTACTGCGATGTCGCTGAACCTATCAGTTCTCGA	2091

RESULT 4
US-08-765-081-1
; Sequence 1, Application US/08765081
; Patent No. 5798260
; GENERAL INFORMATION:
; APPLICANT: Tark, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli O157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible/Pentium
; OPERATING SYSTEM: MS-Windows 3.1
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/765,081
; FILING DATE: March 26, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997


```
/ REFERENCE/DOCKET NUMBER: CHOR-1-10286
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)
/ TELEFAX: 1-206-224-0779
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 8041 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Escherichia coli O157:H7
/ STRAIN: 86-24 NALR
/ IMMEDIATE SOURCE:
/ CLONE: pEAR
/ US-08-765-081-1

Query Match      100.0%; Score 2091; DB 1; Length 8041;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2091; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGAATACCACTCTGCGCTTCGAGTCAATTCCTGTCCTCGGATTTTCAGCCAGCAGC 60
DB 3036 ATGCGAATACCACTCTGCGCTTCGAGTCAATTCCTGTCCTCGGATTTTCAGCCAGCAGC 3095

QY 61 ATAGCTGTCGAGAGGATGATGATGTTGTCCTGCGATCCGGTATGAGAAAAGCTGACT 120
DB 3096 ATAGCTGTCGAGAGGATGATGATGTTGTCCTGCGATCCGGTATGAGAAAAGCTGACT 3155

QY 121 AACGACGCCAGCTGTTCTGATAGTACGAGGAGGATTCAGTCCAGCCAGTACCAC 180
DB 3156 AACGACGCCAGCTGTTCTGATAGTACGAGGAGGATTCAGTCCAGCCAGTACCAC 3215

QY 181 GATCTGCGCGAGGCTCTGAGATCAGTAGAGGTTGTTGATGTTGAAAGTGGTACGGGTAAA 240
DB 3216 GATCTGCGCGAGGCTCTGAGATCAGTAGAGGTTGTTGATGTTGAAAGTGGTACGGGTAAA 3275

QY 241 ACCGAGGGCTGGAATCAGCATCCGAGGATCCGAGGATTCAGCTGATGATGAT 300
DB 3276 ACCGAGGGCTGGAATCAGCATCCGAGGATCCGAGGATTCAGCTGATGATGAT 3335

QY 301 GATGTTGTTGCTCAGGCGCGAAGCAGTACGTTGACTCCCAACGGTTTTTCTGCCATGAT 360
DB 3336 GATGTTGTTGCTCAGGCGCGAAGCAGTACGTTGACTCCCAACGGTTTTTCTGCCATGAT 3395

QY 361 ACCGGGTTCAAGCCCTCTGCGCGCATTTAGCGTTATGAGGTTATCAGGGGCGGATG 420
DB 3396 ACCGGGTTCAAGCCCTCTGCGCGCATTTAGCGTTATGAGGTTATCAGGGGCGGATG 3455

QY 421 TCACACTGTATGGCTCTGATCGATGCGGCGTGGTGAATATCATTACCAAGAAGAT 480
DB 3456 TCACACTGTATGGCTCTGATCGATGCGGCGTGGTGAATATCATTACCAAGAAGAT 3515

QY 481 GCAGACAAATGGCTCTCTTCGTTCAATCAGGGCTGAATCTCAGGAAAGCAAAATGG 540
DB 3516 GCAGACAAATGGCTCTCTTCGTTCAATCAGGGCTGAATCTCAGGAAAGCAAAATGG 3575

QY 541 GGTAAACAGCAGCGGTTTAAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTTCTGCAGC 600
DB 3576 GGTAAACAGCAGCGGTTTAAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTTCTGCAGC 3635

QY 601 CTGACGATAGCGGTAGCACAACAGCGTCAAGGTTTCATCGGTCACATCACTAGCGAT 660
DB 3636 CTGACGATAGCGGTAGCACAACAGCGTCAAGGTTTCATCGGTCACATCACTAGCGAT 3695

QY 661 ACAGCAGGACCGGTTATTCCTTATCCAGGATTCAGAAATTAATCTTGGTGCAGT 720
DB 3696 ACAGCAGGACCGGTTATTCCTTATCCAGGATTCAGAAATTAATCTTGGTGCAGT 3755

QY 721 CTTGACTGGAAGGCGTCGAGCAGGATGTGCTCTGTTTGTATGGATACCAACCCGGCAG 780
```


Db 4836 CAGAAATTATTCGCTACTGAGCGCTGTACAGAGAAAGTGTATGATGAGAAAGAGATAC 4895
Qy 1861 CTGAAGAGCTGAGCGGTGTGATGCGAGCTCTGTCTGTGGAAGATGACCGGATCCCTGACG 1920
Db 4896 CTGAAGAGCTGAGCGGTGTGATGCGAGCTCTGTCTGTGGAAGATGACCGGATCCCTGACG 4955
Qy 1921 CTGAATGCTGCGGTGAATTAACCTGCTCAACAAGGATTACAGTGACGTGAGCGCTGTACAGT 1980
Db 4956 CTGAATGCTGCGGTGAATTAACCTGCTCAACAAGGATTACAGTGACGTGAGCGCTGTACAGT 5015
Qy 1981 GCGGTGAAGAGTACGCTGTATGCGCGGTGATTACTTCCAGACGGGATCATCAACACAGGA 2040
Db 5016 GCGGTGAAGAGTACGCTGTATGCGCGGTGATTACTTCCAGACGGGATCATCAACACAGGA 5075
Qy 2041 TATGTATACCTGAGCGGAATTACTGCGATGTCGCTGAACATATCAGTTCTGA 2091
Db 5076 TATGTATACCTGAGCGGAATTACTGCGATGTCGCTGAACATATCAGTTCTGA 5126

RESULT 5

US-09-098-082-1
; Sequence 1, Application US/09098082
; Patent No. 6040421
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible/Pentium II
; OPERATING SYSTEM: MS-Windows 95
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,082
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/765,081
; FILING DATE: March 26, 1997
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: CHOR-1-12402
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli 0157:H7
; STRAIN: 86-24 NALR
; IMMEDIATE SOURCE:
; CLONE: PEAR
; US-09-098-082-1

Query Match 100.0%; Score 2091; DB 3; Length 8041;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 2091; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGCAATAACCACTCTGGCTTCGGTAGTCATTCCTCTGTCTCGGATTTTCAGCCAGCAGC 60
Db 3036 ATGCGAATAACCACTCTGGCTTCGGTAGTCATTCCTCTGTCTCGGATTTTCAGCCAGCAGC 3095
Qy 61 ATAGCTCTGCGAGGAGTGTATGATGTCTCGSCATCCGGCTATGAGAAAAGACTGACT 120
Db 3096 ATAGCTCTGCGAGGAGTGTATGATGTCTCGSCATCCGGCTATGAGAAAAGACTGACT 3155
Qy 121 AACCGAGCCGCACTGTTCTGTGATTTAGCCAGAGGAAATTGCAGTCCAGCCAGTACAC 180
Db 3156 AACCGAGCCGCACTGTTCTGTGATTTAGCCAGAGGAAATTGCAGTCCAGCCAGTACAC 3215
Qy 181 GATCTGCGGAGGCTCTGAGATCAGTAGAGGGTGTGATGTTGAAAGTGTACGGGTAAA 240
Db 3216 GATCTGCGGAGGCTCTGAGATCAGTAGAGGGTGTGATGTTGAAAGTGTACGGGTAAA 3275
Qy 241 ACCGAGGGCTGGAAATCAGCATCCGAGGAATGCCAGCCAGTTTACAGCTGATACCTGATT 300
Db 3276 ACCGAGGGCTGGAAATCAGCATCCGAGGAATGCCAGCCAGTTTACAGCTGATACCTGATT 3335
Qy 301 GATGCTGTTCTGAGGCGGAGAGCAGTGACTGACTCCCAACGGTTTTTCTGCCATGAAT 360
Db 3336 GATGCTGTTCTGAGGCGGAGAGCAGTGACTGACTCCCAACGGTTTTTCTGCCATGAAT 3395
Qy 361 ACCGGGTTTCATGCCCCCTCTGCGGCCATTGAGGCTATTGAGGTTATCAGGGGCCGATG 420
Db 3396 ACCGGGTTTCATGCCCCCTCTGCGGCCATTGAGGCTATTGAGGTTATCAGGGGCCGATG 3455
Qy 421 TCCACACTGTATGCTCTGATGCGATGGGCGGTGTGTAATATCATTTACAGAAAAGAT 480
Db 3456 TCCACACTGTATGCTCTGATGCGATGGGCGGTGTGTAATATCATTTACAGAAAAGAT 3515
Qy 481 GCAGACAAATGGCTCTCTTCGTCATGCGAGGCTGATCTGCAGGAAAGCAACAAATGG 540
Db 3516 GCAGACAAATGGCTCTCTTCGTCATGCGAGGCTGATCTGCAGGAAAGCAACAAATGG 3575
Qy 541 GGTAAACAGCAGCCAGTTTAATTTCTGGAGCAGTGGTCCCCTTGTGGATGATTTCTGTGACG 600
Db 3576 GGTAAACAGCAGCCAGTTTAATTTCTGGAGCAGTGGTCCCCTTGTGGATGATTTCTGTGACG 3635
Qy 601 CTGAGGTACCGGTAGCAGACAAACAGGTCAGGGTTTCATCGTCACTCATCTAGCGAT 660
Db 3636 CTGAGGTACCGGTAGCAGACAAACAGGTCAGGGTTTCATCGTCACTCATCTAGCGAT 3695
Qy 661 ACAGCAGCAGCGCTATTTCCTTATCCACGAGGTCAACAGAAATTAATCTTTGGTGCAGT 720
Db 3696 ACAGCAGCAGCGCTATTTCCTTATCCACGAGGTCAACAGAAATTAATCTTTGGTGCAGT 3755
Qy 721 CTTGACTGGAAGCGTGGAGCAGGATGCTCTGTTTGTATGATGATACACCCGCGAG 780
Db 3756 CTTGACTGGAAGCGTGGAGCAGGATGCTCTGTTTGTATGATGATACACCCGCGAG 3815
Qy 781 CGTTATGATAACCGGATGGCAACTGGGAGTCTGACGGGGGATGACCGGACCCCTG 840
Db 3816 CGTTATGATAACCGGATGGGCAACTGGGAGTCTGACGGGGGATGACCGGACCCCTG 3875
Qy 841 CGCTATGAGCGAAACAAAATTTTCAGCTGGCTATGATCATCTTTCACTTCGGAACATGG 900
Db 3876 CGCTATGAGCGAAACAAAATTTTCAGCTGGCTATGATCATCTTTCACTTCGGAACATGG 3935
Qy 901 AAATCGTATCTGAATGGAACAGACAGAAATAAGTCTGTGAGCTGTACGAGTGTA 960
Db 3936 AAATCGTATCTGAATGGAACAGACAGAAATAAGTCTGTGAGCTGTACGAGTGTA 3995
Qy 961 CTGAAGCGGACAAATGGGGCTTTCGCGGTGAGCGGGGAGCTTAAGGAATCAACCTT 1020
Db 3996 CTGAAGCGGACAAATGGGGCTTTCGCGGTGAGCGGGGAGCTTAAGGAATCAACCTT 4055
Qy 1021 ATCTGAAATTCATCTGCTTACCTTACCCCTCTGGGAGAAATCTCATCTGTTACCGTGGGGC 1080

Db 4056 ATCTGTAATTCATTACTGCTTACCCCTCTGGAGAACTCATCTGTTACGGTGGGGGC 4115
 Qy 1081 GAGTTTCAGAGCTGTCATGAAAGACGGAGTTGTCTTCCAGACACAGTGAACTTTC 1140
 Db 4116 GAGTTTCAGAGCTGTCATGAAAGACGGAGTTGTCTTCCAGACACAGTGAACTTTC 4175
 Qy 1141 CGGCAGAAAAGCTGGTGGTATTTGCTGAGGATGAGTGGCATCTCACGGATGCACTTGG 1200
 Db 4176 CGGCAGAAAAGCTGGTGGTATTTGCTGAGGATGAGTGGCATCTCACGGATGCACTTGG 4235
 Qy 1201 CTGACTCGGGACCGCTGATGAACATCATGAGCAATTCGGGGACACTTCAGTCCGGCT 1260
 Db 4236 CTGACTCGGGACCGCTGATGAACATCATGAGCAATTCGGGGACACTTCAGTCCGGCT 4295
 Qy 1261 GCATATCTGGTCTGGGATGGCAGATGCCCTGGACGCTGAAGCGGTGGACACGGGA 1320
 Db 4296 GCATATCTGGTCTGGGATGGCAGATGCCCTGGACGCTGAAGCGGTGGACACGGGA 4355
 Qy 1321 TATAAGGCACCCAGAAATGGGACAGCTACATAAAGGGATTAGTGGTGTCTCCGGGACGGGA 1380
 Db 4356 TATAAGGCACCCAGAAATGGGACAGCTACATAAAGGGATTAGTGGTGTCTCCGGGACGGGA 4415
 Qy 1381 AAAACAATCTACTTGTGTAACCCCGACCTGAAGCCGGAAGAGAGCGTCAGTTATGAGCT 1440
 Db 4416 AAAACAATCTACTTGTGTAACCCCGACCTGAAGCCGGAAGAGAGCGTCAGTTATGAGCT 4475
 Qy 1441 GGGGTGTATTACGATAACCCCGCGCTCTGAATGCCAATGTACAGGTTTTATGACTGAC 1500
 Db 4476 GGGGTGTATTACGATAACCCCGCGCTCTGAATGCCAATGTACAGGTTTTATGACTGAC 4535
 Qy 1501 TTCTCCAAAGATGTCCTTATTCATTAATGAATGAACCAATAGCTATGTAACAGC 1560
 Db 4536 TTCTCCAAAGATGTCCTTATTCATTAATGAATGAACCAATAGCTATGTAACAGC 4595
 Qy 1561 GGAAGCCCGGTGACGGTGTGGAAATTTCCGGGCACATTCGGCTGTGCTCAGAGGAT 1620
 Db 4596 GGAAGCCCGGTGACGGTGTGGAAATTTCCGGGCACATTCGGCTGTGCTCAGAGGAT 4655
 Qy 1621 GTACAGTGTACTGAATTAACCTGACCCGAGTGAACAAAGTGAACAAAGTGTGTAACAA 1680
 Db 4656 GTACAGTGTACTGAATTAACCTGACCCGAGTGAACAAAGTGAACAAAGTGTGTAACAA 4715
 Qy 1681 GGTCCGCGCTGAGTTATACCTGACACATGGTGAATCGGAACCTGAACTGGCAGATC 1740
 Db 4716 GGTCCGCGCTGAGTTATACCTGACACATGGTGAATCGGAACCTGAACTGGCAGATC 4775
 Qy 1741 ACCGAAGAGTGGCATCATGGCTGGTCCCGTTATCGCGGGAACACACACCGTTTCAAC 1800
 Db 4776 ACCGAAGAGTGGCATCATGGCTGGTCCCGTTATCGCGGGAACACACACCGTTTCAAC 4835
 Qy 1801 CAGAAATTTCTGTCACCTGAGGCTGTACAGAAAGTGTATGATGAGAAAGAGATAC 1860
 Db 4836 CAGAAATTTCTGTCACCTGAGGCTGTACAGAAAGTGTATGATGAGAAAGAGATAC 4895
 Qy 1861 CTGAAGCCTGGACCGTGGTGGATGACAGTCTGTCTGGAAGATGACGATGCCCTGACG 1920
 Db 4896 CTGAAGCCTGGACCGTGGTGGATGACAGTCTGTCTGGAAGATGACGATGCCCTGACG 4955
 Qy 1921 CTGAATGCTGGTGGTAATACCTGCTCAAAGAGATGACGATGAGTGGTGGTGGTGGT 1980
 Db 4956 CTGAATGCTGGTGGTAATACCTGCTCAAAGAGATGACGATGAGTGGTGGTGGTGGT 5015
 Qy 1981 GCCGTAAGAGTACGCTGTATGCGGTGATTCTTCCAGACGGGATCATCAACACAGGA 2040
 Db 5016 GCCGTAAGAGTACGCTGTATGCGGTGATTCTTCCAGACGGGATCATCAACACAGGA 5075
 Qy 2041 TATGTGATACCTGAGCGAAATTAAGTGTGATGCTGAACTATCAGTTCTGA 2091
 Db 5076 TATGTGATACCTGAGCGAAATTAAGTGTGATGCTGAACTATCAGTTCTGA 5126

; Sequence 1, Application PC/TUS9506994
 ; GENERAL INFORMATION:
 ; APPLICANT: Children's Hospital & Medical Center
 ; APPLICANT: University of Washington
 ; APPLICANT: Washington State University Research Foundation
 ; APPLICANT: TARR, PHILLIP I
 ; APPLICANT: BILGE, SIMA S
 ; APPLICANT: BESSER, THOMAS E
 ; APPLICANT: VARY JR, JAMES C
 ; TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
 ; STREET: SUITE 2800, 1420 FIFTH AVENUE
 ; CITY: SEATTLE
 ; STATE: WASHINGTON
 ; COUNTRY: USA
 ; ZIP: WA 98101
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/06994
 ; FILING DATE: 07-JUN-95
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/265,714
 ; FILING DATE: 24-JUN-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BRODERICK, THOMAS F
 ; REGISTRATION NUMBER: 31,332
 ; REFERENCE/DOCKET NUMBER: CHOR-18591
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 682 8100
 ; TELEFAX: (206) 224 0779
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 8041 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Escherichia coli O157:H7
 ; STRAIN: 86-24 NALR
 ; IMMEDIATE SOURCE:
 ; CLONE: PEAR
 ; PCT-US95-06994-1

Query Match 100.0%; Score 2091; DB 5; Length 8041;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2091; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ATCGGAATAACCACTCTGGCTTCGGTAGTCATTCCTCGGTCTCGGATTTTCAGCCAGCAGC 60
 Db 3036 ATCGGAATAACCACTCTGGCTTCGGTAGTCATTCCTCGGTCTCGGATTTTCAGCCAGCAGC 3095
 Qy 61 ATAGTGTCTGCAGAGGATGTGATGATTGTCTCGGCATCCGGCTATGAGAAAAAGCTGACT 120
 Db 3096 ATAGTGTCTGCAGAGGATGTGATGATTGTCTCGGCATCCGGCTATGAGAAAAAGCTGACT 3155
 Qy 121 AACGAGCCGCGAGTGTTCGTGATTAGCCAGGAGAAATTCAGTCCAGCCAGTACAC 180
 Db 3156 AACGAGCCGCGAGTGTTCGTGATTAGCCAGGAGAAATTCAGTCCAGCCAGTACAC 3215
 Qy 181 CATCTGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGTGACGGSTAAA 240
 Db 3216 CATCTGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGTGACGGSTAAA 3275

QY 241 ACCGAGGGCTGGAATCAGCATCCGAGGAATCCGAGCACTTACACGCTGATCTGATT 300
 Db 3276 ACCGAGGGCTGGAATCAGCATCCGAGGAATCCGAGCACTTACACGCTGATCTGATT 3335
 QY 301 GATGTTGTTCTGTCAGGGCGGAAGCAGTGAAGTGAATCCCAACGGTTTTCGCGCATGAAT 360
 Db 3336 GATGTTGTTCTGTCAGGGCGGAAGCAGTGAAGTGAATCCCAACGGTTTTCGCGCATGAAT 3395
 QY 361 ACCGGTTTCAATGCGCCCTCTGCGCCGCAATGAGCGGTATTGAGGTTATCAGGGGGCGGATG 420
 Db 3396 ACCGGTTTCAATGCGCCCTCTGCGCCGCAATGAGCGGTATTGAGGTTATCAGGGGGCGGATG 3455
 QY 421 TCACACTGTATGGCTCTGATCGGATGGCGGTGGTGGTGAATATCATATCCAGAAAGAAAT 480
 Db 3456 TCACACTGTATGGCTCTGATCGGATGGCGGTGGTGGTGAATATCATATCCAGAAAGAAAT 3515
 QY 481 CGAGCAAAATGGCTCTTCCGCTCAATGAGCGGTGAATCTCGAGGAAGCAACAAATGG 540
 Db 3516 CGAGCAAAATGGCTCTTCCGCTCAATGAGCGGTGAATCTCGAGGAAGCAACAAATGG 3575
 QY 541 GGTAAACAGAGCGAGTTTAAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTTCTCTCAGC 600
 Db 3576 GGTAAACAGAGCGAGTTTAAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTTCTCTCAGC 3635
 QY 601 CTGAGGTACGGGTAGCACAACAGCGTCAAGGTTCATCGGTACATCAGTGAAGGAT 660
 Db 3636 CTGAGGTACGGGTAGCACAACAGCGTCAAGGTTCATCGGTACATCAGTGAAGGAT 3695
 QY 661 ACAGAGGACCGGTTATCTTATCCAGCGAGTACAGAAATTAATTTCTGGTCAAGT 720
 Db 3696 ACAGAGGACCGGTTATCTTATCCAGCGAGTACAGAAATTAATTTCTGGTCAAGT 3755
 QY 721 CTGAGTGAAGCGTCGAGCAGGATGTCTCTGGTTGATATGATACCAACCGGGAG 780
 Db 3756 CTGAGTGAAGCGTCGAGCAGGATGTCTCTGGTTGATATGATACCAACCGGGAG 3815
 QY 781 CGTTATGATTAACCGGATCGGCAACTGGGAGTCTGACGGGGGATATCAACCGGACCTG 840
 Db 3816 CGTTATGATTAACCGGATCGGCAACTGGGAGTCTGACGGGGGATATCAACCGGACCTG 3875
 QY 841 CGCTATGAGCGAAACAAAATTTTCACTGGCTATGATCATATTTTCACTTTCGGAACATGG 900
 Db 3876 CGCTATGAGCGAAACAAAATTTTCACTGGCTATGATCATATTTTCACTTTCGGAACATGG 3935
 QY 901 AAATCGTATCTGAATCTGAGCAGCAGCAAGAAATTAAGGTCGTGAGCTGTACGCAAGTGA 960
 Db 3936 AAATCGTATCTGAATCTGAGCAGCAGCAAGAAATTAAGGTCGTGAGCTGTACGCAAGTGA 3995
 QY 961 CTGAAGCGGACAAAATGGGGGCTTGGCGGTCAAGCGGGGAGCTTAAGGAATCGAACTTT 1020
 Db 3996 CTGAAGCGGACAAAATGGGGGCTTGGCGGTCAAGCGGGGAGCTTAAGGAATCGAACTTT 4055
 QY 1021 ATCTGAATTCATTAATCTGTTACCTCTGGAGGAATCTCATCTGGTTACGGTGGGGGC 1080
 Db 4056 ATCTGAATTCATTAATCTGTTACCTCTGGAGGAATCTCATCTGGTTACGGTGGGGGC 4115
 QY 1081 GAGTTTCAGAGCTCGTCCATGAAGACGGAGTTGCTTTGCGAGCAGCAGTGAATCTTTC 1140
 Db 4116 GAGTTTCAGAGCTCGTCCATGAAGACGGAGTTGCTTTGCGAGCAGCAGTGAATCTTTC 4175
 QY 1141 CGGAGAAAAGCTGGTGGTATTTGCTGAGGATGAGTGCATCTCAACGATCACTTGG 1200
 Db 4176 CGGAGAAAAGCTGGTGGTATTTGCTGAGGATGAGTGCATCTCAACGATCACTTGG 4235
 QY 1201 CTGACTGCGGGCAGCGCTATGAACATCATGAGCAATTCGGGGGACACTTCAGTCCGGT 1260
 Db 4236 CTGACTGCGGGCAGCGCTATGAACATCATGAGCAATTCGGGGGACACTTCAGTCCGGT 4295
 QY 1261 GCATATCTGGTCTGGGATGAGCAGATGCTGAGCAGCTGAAGCGCGGTGACCAACGGGA 1320
 Db 4296 GCATATCTGGTCTGGGATGAGCAGATGCTGAGCAGCTGAAGCGCGGTGACCAACGGGA 4355
 QY 1321 TATAAGGCCACCGAATGGGGCAGCTACATAAAGGAATAGTGGTGTGTCGGGGCAGGGA 1380

Db 4356 TATAAGGCCACCGAATGGGGCAGCTACATAAAGGAATAGTGGTGTGTCGGGCAGGGA 4415
 QY 1381 AAACCAATCTACTTGTAAACCCGACCTGAAGCCGGAAGAGAGCGTCAGTATGAGGCT 1440
 Db 4416 AAACCAATCTACTTGTAAACCCGACCTGAAGCCGGAAGAGAGCGTCAGTATGAGGCT 4475
 QY 1441 GGGTGTATTACGATAAACCCCGCGGTCTGAATGCCAATGTACAGGTTTATGACTGAC 1500
 Db 4476 GGGTGTATTACGATAAACCCCGCGGTCTGAATGCCAATGTACAGGTTTATGACTGAC 4535
 QY 1501 TTCTCCAAACAGATGTCTCTTATCCATAAATGATAACCAACCAATAGCTATGTAACAGC 1560
 Db 4536 TTCTCCAAACAGATGTCTCTTATCCATAAATGATAACCAACCAATAGCTATGTAACAGC 4595
 QY 1561 GGAAGGCCCGGTTCGACGGTGGGAATTTCCCGGCACATTCGCGGTGTGGTCAAGGAT 1620
 Db 4596 GGAAGGCCCGGTTCGACGGTGGGAATTTCCCGGCACATTCGCGGTGTGGTCAAGGAT 4655
 QY 1621 GTACCGTGTCACTGAATTAACCTGACCCGGAAGTGAACAACTGATGATGATAACAAA 1680
 Db 4656 GTACCGTGTCACTGAATTAACCTGACCCGGAAGTGAACAACTGATGATGATAACAAA 4715
 QY 1681 GGTCCGCGCTGAGTTATACCCCTGAACACATGGTGAATGCGAAACTGAACTGCGAGATC 1740
 Db 4716 GGTCCGCGCTGAGTTATACCCCTGAACACATGGTGAATGCGAAACTGAACTGCGAGATC 4775
 QY 1741 ACCGAGAGGTGGCATCATGGTGGGTGGCGGTTCATCGCGGGAACAAACACCACTTTCACC 1800
 Db 4776 ACCGAGAGGTGGCATCATGGTGGGTGGCGGTTCATCGCGGGAACAAACACCACTTTCACC 4835
 QY 1801 CAGAAATTTATCTGACAGCGGTGTACAGAAAGTGTATGATGAAAGGAGAGATAC 1860
 Db 4836 CAGAAATTTATCTGACAGCGGTGTACAGAAAGTGTATGATGAAAGGAGAGATAC 4895
 QY 1861 CTGAAGCGCTGGAGCGGTGGTGGATGCAAGTCTGCTGGGAAGATGACGGATGCCCTGACG 1920
 Db 4896 CTGAAGCGCTGGAGCGGTGGTGGATGCAAGTCTGCTGGGAAGATGACGGATGCCCTGACG 4955
 QY 1921 CTGAATGTCGGGTGAATTAACCTGCTCAACAGAGTTACAGTGAAGTGTGATGAGAAAGGAGATAC 1980
 Db 4956 CTGAATGTCGGGTGAATTAACCTGCTCAACAGAGTTACAGTGAAGTGTGATGAGAAAGGAGATAC 5015
 QY 1981 GCGGTGAAGTACGCTGTATGCGCGGTGATTTACTTCCAGACGGGATCATCAACCAACAGGA 2040
 Db 5016 GCGGTGAAGTACGCTGTATGCGCGGTGATTTACTTCCAGACGGGATCATCAACCAACAGGA 5075
 QY 2041 TATGTGATCTGAGCGAAATTAATCTGATGTGCTGAACTATCAGTTCTGA 2091
 Db 5076 TATGTGATCTGAGCGAAATTAATCTGATGTGCTGAACTATCAGTTCTGA 5126

RESULT 7

US-09-453-702B-57/c
 ; Sequence 57, Application US/09453702B
 ; Patent No. 6365723

GENERAL INFORMATION:

APPLICANT: Blattner, Frederick R.
 ; Burland, Valerie
 ; Ferna, Nicole T.
 ; Plunkett, Guy
 ; Welch, Rod
 ;

TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady
 ; STREET: 1 South Pinckney Street
 ; CITY: Madison
 ; STATE: WI
 ; COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage

QY 1489 TTTATGACTGCTTCCAAAGAGTTGTCTTATTTCCTATAATGATAAC 1539
 DB 1492 TTTAATACAGATTAAAGATAAGCTTACCAACTACGATACCGGTGAATC 1542

RESULT 9
 US-09-489-039A-962
 ; Sequence 962, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; FILE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 962
 ; LENGTH: 2100
 ; TYPE: DNA
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-962

Query Match 5.0%; Score 105.4; DB 4; Length 2100;
 Best Local Similarity 51.9%; Pred. No. 2.1e-24;
 Matches 238; Conservative 0; Mismatches 221; Indels 0; Gaps 0;
 QY 1060 CATCTGCTTACGCTGGGGGGGAGTTTCAGAGCTCGTCCATGAAGACGAGTTGTCCTT 1119
 DB 1108 CAGCTGCTGCTTCGCGGCGGATGCGCCACGACAACTGAAGATCCGCTCAACTG 1167
 QY 1120 GCGAGCAGGTGAACCTTCCGGCAGAAAGCTGCTGGTATTTCGTAGGATGATGG 1179
 DB 1168 AGCAGCGCGGCCAGTCAACGCTCGGCTAGCCAGTACGCCCTGTTTATCGAAGACGAATG 1227
 QY 1180 CATCTCAGGATGACATTCGCTGACTCGCGGAGCGCTATCAACATCATAGCAATTC 1239
 DB 1228 CGCATATCAGCGCTCGCGCTGACCCAGCGATTCGATGACACCAATCAGACTAT 1287
 QY 1240 GGGGACACTTCAGTCCGCTGCATATCTGCTGGGATGTCGACATGCTGGACGCTG 1299
 DB 1288 GCGCATCACTGGAGCCCGCGCTATCTGCTGTATTAACGCCACCGATACCGTCACTG 1347
 QY 1300 AAAGCGGTGTACCCCGGATATAGGACCCAGATGGGCGAGCTACATAAAGGAT 1359
 DB 1348 AAAGCGGCTGGGCGAGCGGCTTAAAGCCCGCTGCTGCTGAGCTTAAACCCGCTGG 1407
 QY 1360 AGTGGTGTGTCGGGCGAGGAAACAAATCTACTTGTAAACCCCGACTGAAAGCGGAA 1419
 DB 1408 ACCACCACTCTGTCGCGCTCGTGACGATCTGCTGTAACCCGATCTGAAACCGGAA 1467
 QY 1420 GAGAGCTCAGTTATGAGCTGGGTGTTATAGATAACCCCGCGCTGATGCAAT 1479
 DB 1468 ACCAGCGAAGCTTCGAGCTCGTCTCTACTACCGGCGGAGAGGCTGGCTTGAAT 1527
 QY 1480 GTACAGCTTTATGACTGACTTCTCCAAAGATTGTC 1518
 DB 1528 GTCAAGGACGATCACCACTCCAGATAATGTGAC 1566

RESULT 10
 US-08-480-510-1
 ; Sequence 1, Application US/08480510
 ; Patent No. 5747028
 ; GENERAL INFORMATION:
 ; APPLICANT: Calderwood, Stephen B.
 ; APPLICANT: Buterton, Joan R.
 ; APPLICANT: Mekalanos, John J.
 ; TITLE OF INVENTION: HETEROLOGOUS ANTIGENS IN LIVE
 ; TITLE OF INVENTION: CELL
 ; TITLE OF INVENTION: VACCINE STRAINS

NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 COMPUTER: IBM PS/2 Model 50Z or 55SX
 OPERATING SYSTEM: MS-DOS (Version 5.0)
 SOFTWARE: WordPerfect (Version 5.1)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/480,510
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/020,501
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Clark, Paul T.
 REGISTRATION NUMBER: 30,162
 REFERENCE/DOCKET NUMBER: 00786/136001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 542-5070
 TELEFAX: (617) 542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1535
 TYPE: nucleic acid
 STRANDEDNESS: both
 TOPOLOGY: linear
 US-08-480-510-1

Query Match 5.0%; Score 105; DB 1; Length 1535;
 Best Local Similarity 57.4%; Pred. No. 2.3e-24;
 Matches 214; Conservative 0; Mismatches 150; Indels 9; Gaps 1;
 QY 71 CAGAGATGTGATGATGCTCGGCATCCGCTATGAGAAAAGCTGACTAACGACCG 130
 DB 1168 CGGATGAACCATGCTGCTCACTCGCGGGGATACGCGCAAGTATTCAAAATGCAACG 1227
 QY 131 CCAGTGTTCCTGATTAGCCAGGAGAAATGCACTCCAGCCAGTACCCAGATCTGCGG 190
 DB 1228 CCAGTATCAGTGTGATTTCAAGAGAAGATCTGGATCTCGCTATTACCGTATGTGACG 1287
 QY 191 AGGCTCTGATCTAGTACGAGGTGCTGATGTTGAAAGTGTGACGGTAAACCGAGGGC 250
 DB 1288 ATGCGCTAAAAGGCTACCGGCTGTGACAGTCAACGGAGGGGGGATACCTACC----- 1340
 QY 251 TGGAAATCAGCATCCGAGGATCCAGCAGTTTACAGCTGATACCTGATTGATGCTGTC 310
 DB 1341 --GATATCAGATTCCTGCTGATGGATCAACTATCTTAICTTCTGCTGATGTTAAGC 1398
 QY 311 GTCAGGGCGGAAGCAGTACGTACTCCCAACCGTTTTCTGCGCATGAATACCGGGTCA 370
 DB 1399 GCCAAACCTCACCGCAGACCGCTCCAAACAGCGATGTCGCGGGGCAATGAGCAAGTTGGT 1458
 QY 371 TGCCCCCTTGGCCCATTCAGGCTATTGAGGTTATCAGGGGCGCATGTCACACTGT 430
 DB 1459 TACCGCACCTCAAGCGATTGACGTATCGAGGTGATCCGTGGGCCCGATGTCTACGCTGT 1518
 QY 431 ATGGCTCTGATGC 443
 DB 1519 ACGGCTCGGATGC 1531

RESULT 11
 PCT-US94-01780-1
 ; Sequence 1, Application PC/TUS9401780
 ; GENERAL INFORMATION:

APPLICANT: Calderwood, Stephen B.
APPLICANT: Butterton, Joan R.
APPLICANT: Mekalanos, John J.
TITLE OF INVENTION: HETEROLOGOUS ANTIGENS IN LIVE CELL
TITLE OF INVENTION: VACCINE STRAINS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01780
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/020,501
FILING DATE: February 22, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/136001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1535
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
PCT-US94-01780-1

Query Match 5.0%; Score 105; DB 5; Length 1535;
Best Local Similarity 57.4%; Pred. No. 2.3e-24;
Matches 214; Conservative 0; Mismatches 150; Indels 9; Gaps 1;
QY 71 CAGAGGATGTGATGTTGTCGGCATCCGGTATGAGAAAAAGTGACTAACGAGCGG 130
Db 1168 CGATGAACCATGGTGGTCACTGCGGGGATACGCGCAAGTGATTCAAAATGCACAG 1227
QY 131 CCAGTGTTCGTGATTACCGCAGGAGGAATTGAGTCCAGCCAGTACACGATCTGGCGG 190
Db 1228 CCAGTATCAGTGTGATTTCAAGAGAGATCTGGAATCTCGCTATTACCGTGATGACCG 1287
QY 191 AGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGMAAGTGGTACGGGTAAACCGGAGGCG 250
Db 1288 ATGCGGTAAACCGTACCGGGTGTGACAGTACCGGAGGGGGCGNACTACC----- 1340
QY 251 TGAATATCAGATCCAGAGGATGCCAGCCAGTTACAGCTGATATGATGATGTTGTC 310
Db 1341 --GATATCAGATTCGTGGTATGGGATCAAACTATACTCTTATCTTGGTGGATGTAAGC 1398
QY 311 GTCAGGGCGGAGCAGTACGTCAGTCCCAAGGTTTCTGCCATGAATCCGGTTCA 370
Db 1399 GCCAAACCTCAGCCAGACCCGTCCAAACAGCGATGGCCCGGCGATTTGCAAGTTGGT 1458
QY 371 TGCCCTCTGGCCGATTTAGCGTATTGAGGTTATCAGGGGGCGGATGTCACACTGT 430
Db 1459 TACCGCACTGCAAGCGATTGAACGATCGAGTGATCGTGGCCCGGATGTCACCTGT 1518
QY 431 ATGGCTCTGATGC 443
Db 1519 ACGGCTCGATGC 1531

RESULT 12
US-09-543-681A-3764
; Sequence 3764, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 3764
; LENGTH: 2034
; TYPE: DNA
; ORGANISM: Proteus mirabilis
; US-09-543-681A-3764

Query Match 4.6%; Score 96.8; DB 4; Length 2034;
Best Local Similarity 50.0%; Pred. No. 1.6e-21;
Matches 276; Conservative 0; Mismatches 267; Indels 9; Gaps 1;
QY 63 AGCTGCTGCAGAGATGTGATGATTGTCTCGGCATCCGGCTATCAGAAAAAGCTGACTAA 122
Db 108 AGAAGATCCTGAAAGGTTACTGGTGACAAACCGCTTCGGGATTTAAACAGACTGTTGAAGA 167
QY 123 CGAGCGCCAGTGTCTGTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCACGA 182
Db 168 TGCTCCAGCTCTGTTTCAGTTGTAAGTCTGTAACAATAGAACAAATCATATCGTGA 227
QY 183 TCTGGCGGAGCTCTGAGATCAGTAGAGGTTGGATGTTGAAAGTGTGCGGTAAC 242
Db 228 TGTGACTGATCGCTAAAGATGTTCCGGGGTGTCTAGT-----GACGGGGGGCGG 278
QY 243 CGGAGGCTGGAATCAGCATCCGAGGATCCAGGATCCAGCTGATGATGTTGA 302
Db 279 TAGCAGCTCTGATATTAGTATTCTGGTATGGAATCCTTAAGTATACCATGATCTTAGTCGA 338
QY 303 TGGTGTTCGTGAGGGCGGAAGCAGTGCAGTCCCAACCGTTCCTCCATGAATAC 362
Db 339 TGGTAAAGCTGTGCTCATCTCGTGAACCGCTCTAATAGTAACTCAGGTATTGAACA 398
QY 363 CGGTTTCATGCCCTCTGCGCCATGAGCGTATTGAGGTTATCAGGGGGCGGATGC 422
Db 399 AGGTTGGTTACCTCCTTTACCCGCGATTGAGCGTATTGAGGTTAGTAAGAGGCCCAATGTC 459
QY 423 CACACTGTATGGTCTCATGCGATGGCGGTGTGATGATATCATATACCAAGAAAGATGC 482
Db 459 TCTTTATATGGCTCTGATGCCATGGAGCGTTATCAATATTATTACGCGTAAAGCGCA 518
QY 483 AGACAAATGGCTCTTCTTCGCTCAATGAGGGCTGAATCTCGAGGAAGCAAAATGGGG 542
Db 519 AAAAGATGGAACCTTAGTTTACGTGGTGACACAAACGTTGACTGAGCGTAAAAATGAAGG 578
QY 543 TAAACAGCAGCCGTTTAAATTTCTGGACAGTGGTCCCTTGTGGATGATCTGTGACGCT 602
Db 579 AAATACAGGCAAGTAGTTTCTTANGCTGCAAGTCCATGATTGATGATGTTGTTAGTCT 638
QY 603 GCAGGTACGGGG 614
Db 639 AAAGCTGCAAGG 650

RESULT 13
US-09-489-039A-703
; Sequence 703, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2034001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 703

; LENGTH: 2280

; TYPE: DNA

; ORGANISM: Klebsiella pneumoniae

; US-09-489-039A-703

Query Match 3.3%; Score 68.8; DB 4; Length 2280;

Best Local Similarity 57.4%; Pred. No. 4.2e-12;

Matches 124; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 389 TTGAGCGTATGAGGTTATCAGGGGCGCGATGCCACACTGTATGCTCTGTATGCGATGG 448

Db 524 TCGAACGGATTGAAGTTCTCCGCGGTCCGCGAGCGCGCTATGTTCCGCGCGCGCG 583

QY 449 GCGGTGTGGTGAATCATCATTACAGAAAGATCGACAAATGGCTCTTCCGTCAATG 508

Db 584 GCGAGTGGTCAATATCATTACCAACAGCTCCGACCACTTGGCAGGTTGCTGTCTT 643

QY 509 CAGGGCTGAATCTGCAGGAAGCAACAATGGGTAAACAGCCAGTTTAATTTCTGGA 568

Db 644 TCTTACCAACACCGCGGAAACCAACAAGAGCGACCAATCGCGCTAACTTCAATC 703

QY 569 GCAGTGTCCCTTGTGGATGATCTGTACGCCCTGC 604

Db 704 TCAGCGGCCACTGGCGCGGAGGCGCTGACGATGC 739

RESULT 14

US-09-668-113A-1

; Sequence 1, Application US/09668113A

; Patent No. 6410703

; GENERAL INFORMATION:

; APPLICANT: Russo, Thomas A.

; TITLE OF INVENTION: Identification of A Vaccine Candidate from an

; FILE REFERENCE: 11520.0214

; CURRENT APPLICATION NUMBER: US/09/668,113A

; CURRENT FILING DATE: 2000-09-22

; NUMBER OF SEQ ID NOS: 10

; SEQ ID NO 1

; LENGTH: 2495

; TYPE: DNA

; ORGANISM: Escherichia coli

; FEATURE:

US-09-668-113A-1

Query Match 3.3%; Score 68.6; DB 4; Length 2495;

Best Local Similarity 55.1%; Pred. No. 5.2e-12;

Matches 134; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 376 CCTCTGCGCCCATTTGAGGTTATGAGGTTATCAGGGGCGCGATGCCACACTGTATGCC 435

Db 732 CCACCGGAACAGGTTGAGCGTATTGAAGTATCCGCGCCCTTCCGCGCGCGCTACGGT 791

QY 436 TCTGATGCGATGGCGGCTGTGGTGAATCATCATTACAGAAAGATCGACAAATGGCTC 495

Db 792 TCGGGGCGCGCGGGGGTGGTGAATCATCATTACCAACGTCCTCCCAACACTGGCAC 851

QY 496 TCTTCGCTCAATGCAGGCTGAAATCTGCAGGAAGCAACAATGGGGTAAACAGCAGCCAG 555

Db 852 GGTTCGCTCTCGTTATACACCAACAGCGGAAAGTAGCGAAGAGGGGCGCTACGGCTGC 911

QY 556 TTAAATTTCTGGAGCAGTCCCTTGTGGATGATCTGTACGCTCGAGGTACGCGGT 615

Db 912 GCCAATTCAGCTTAGTGGGCTCTGGCTGTGGATGCTCTTACCACGGTTTGTATGTT 971

QY 616 AGC 618

Db 972 AAC 974

RESULT 15

US-09-489-039A-6223

; Sequence 6223, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 6223

; LENGTH: 2388

; TYPE: DNA

; ORGANISM: Klebsiella pneumoniae

; US-09-489-039A-6223

Query Match 3.0%; Score 63.4; DB 4; Length 2388;

Best Local Similarity 54.5%; Pred. No. 2.8e-10;

Matches 127; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 376 CCTCTGCGCCCATTTGAGGTTATGAGGTTATCAGGGGCGCGATGCCACACTGTATGCC 435

Db 583 CCGCGGAGATGATCGAACGTTATCGAAGTATTTCGGCGCGCGCGCGCTACGCC 642

QY 436 TCTGATGCGATGGCGGTGTGGTGAATCATCATTACAGAAAGATCGACAAATGGCTC 495

Db 543 AACGGCGCGCGCGCGGTGTGGTGAATCATCATTACCAAAAAACCGCGGTATGGTGCAC 702

QY 496 TCTTCGCTCAATGAGGCTGAATCTTCAGGAAGCAACAATGGGGTAAACAGCAGCCAG 555

Db 703 GGCCTCATGGAACACCTATATGAACGCGCGCGAGCAAGGATGAAGGCTCCACCAACGC 762

QY 556 TTAAATTTCTGGAGCAGTGTCCCTTGTGGATGATTTCTGTACGCTTCGAGGT 608

Db 763 ACTAACTTCAGCTCAGCGCCCACTGGCGCGCGATTTTAGCTTCGCGCTGTT 815

Search completed: October 12, 2004, 15:58:47

Job time : 161 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2004, 11:56:38 ; Search time 819 Seconds
(without alignments)
10846.132 Million cell updates/sec

Title: US-10-625-972-4

Perfect score: 2091

Sequence: 1 atggaataaacactctggc.....cgctgaactatcagttctga 2091

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 29Jan04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002s:*

7: Geneseqn2003as:*

8: Geneseqn2003bs:*

9: Geneseqn2003cs:*

10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2091	100.0	8041	2	AAT10105
C 2	2089.4	99.9	87563	8	Adc19044 E. coli 0
C 3	2087.8	99.8	86248	9	Adc00087 Enterocae
C 4	2079.8	99.5	76804	6	Abs78942 E. coli C
5	149.4	7.1	1980	7	Acif67549 Phototab
6	149.4	7.1	110000	7	Acif65383 Phototab
7	149.4	7.1	110000	7	Continuation (2 of
8	130.6	6.2	10353	4	Aas46270 DNA encod
9	125.6	6.0	2046	4	Aba89269 Escherich
10	125.6	6.0	12943	4	Aba89255 Escherich
11	105	5.0	1535	2	Aaq1387 Vibrio ch
12	104	5.0	1530	2	Aaq26542 Truncated
13	69.4	3.3	3675	5	Aas80035 DNA encod
14	68.6	3.3	2495	4	Aad03054 ironNec ge
15	66.6	3.2	4039	3	Aaa15184 DNA encod
16	64.2	3.1	758	4	Abk43522 DNA encod
C 17	64.2	3.1	822	4	Aas27059 cDNA enco
C 18	64.2	3.1	822	9	Adb93237 Human CDN
C 19	64.2	3.1	10731	4	Aas46247 DNA encod
20	63.2	3.0	2727	5	Aas93059 DNA encod
21	63.2	3.0	2808	5	Aas88838 DNA encod
22	57.6	2.8	686	4	Aas27475 cDNA enco
23	57.6	2.8	686	4	Abk43834 DNA encod

24	57.6	2.8	686	9	ADB93653
25	56.4	2.7	2790	8	ADA29178
26	52.2	2.5	2000	7	ADA71938
27	51.8	2.5	296	6	ABN75479
28	51.6	2.5	1845	3	AAZ38967
C 29	51.4	2.5	856	6	ABQ29448
30	51.4	2.5	856	6	ABQ29449
31	50.8	2.4	687	6	ABQ36380
C 32	50.8	2.4	687	6	ABQ36381
C 33	50.4	2.4	687	6	ABQ36379
34	50.4	2.4	687	6	ABQ36379
C 35	50.2	2.4	856	6	ABQ29447
36	49.2	2.4	856	6	ABQ29446
37	48.6	2.4	247	6	ABN26822
38	48.6	2.3	1963	4	ABA89976
C 39	48.6	2.3	12264	4	ABA89966
40	48.6	2.3	23654	6	ABS78844
41	48.4	2.3	1844	3	AAZ38968
42	47.8	2.3	885	5	AAF26439
43	47	2.2	2544	2	AAX91730
44	47	2.2	2634	2	AAX91603
45	47	2.2	12677	4	AAS46236

ALIGNMENTS

RESULT 1

AAT10105
ID AAT10105 standard; DNA; 8041 BP.

AC AAT10105;

DT 13-MAY-1996 (first entry)

DE Adherence conferring plasmid pear.

KW adhesin; plasmid pear; vector; vaccine; intestine colonisation; ds.

OS Escherichia; coli.

OS Synthetic.

OS Chimeric.

Key

Location/Qualifiers

primer_bind complement(2867..2890)

/tag= a /note= "primer for E.coli O157:H7 DNA"

primer_bind complement(2895..2914)

/tag= b /note= "primer for pear"

CDS 3036..5126

/tag= c

/product= "adhesin"

/trans_except= pos:3912..3914; aa:Gly

/note= "claim 1, page 36"

misc_feature 3271..3310

/tag= d /note= "region of transposon TnpHoA insertion"

misc_feature 3801..3840

/tag= e /note= "region of transposon TnpHoA insertion"

primer_bind complement(5159..5183)

/tag= g /note= "primer for E.coli O157:H7 DNA"

primer_bind complement(5176..5196)

/tag= f /note= "primer for pear"

CDS complement(6449..7024)

/tag= h /note= "terE homologue"

CDS complement(7092..7670)

/tag= i /note= "terD homologue"

FT

XX WO9600233-A1.
PN
XX
XX
PD 04-JAN-1996.
XX
XX PF 07-JUN-1995; 95WO-US006994.
XX
XX PR 24-JUN-1994; 94US-00265714.
XX
XX (CHIL-) CHILDREN'S HOSPITAL & MEDICAL CENT.
PA (UNIW) UNIV WASHINGTON.
PA (UNIW) UNIV WASHINGTON STATE RES FOUND.
XX
XX Tarr PI, Bilge SS, Besser TE, Vary JC;
XX
XX WPI; 1996-068826/07.
DR P-PSDB; AAR75366.
XX
XX Chromosomal DNA from E. coli O157:H7 encoding epithelial adhesin -
PT isolated on plasmid pSC (overlap), for use as a vaccine to mediate
PT bacterial colonisation of bovine intestine.
XX
XX Disclosure; Page 23-30; 42pp; English.
XX
XX Adherence conferring plasmid pear (AAT10105) comprises Escherichia coli
CC O157:H7 chromosomal DNA plus the Stratagene SK+ vector. It includes 3
CC open reading frames, 2 of which are homologues of tere and tereD genes
CC necessary for tellurite resistance. The third ORF is homologous to the
CC IrgA gene and encodes an adhesin (AAR75366) that enables E. coli O157:H7,
CC an antibiotic-resistant, virulent and common food-borne pathogen, to
CC adhere to epithelial cells. This ORF can be utilised in the prodn. of
CC adhesin for use as a vaccine to prevent disease or colonisation of
CC mucosal surfaces by O157:H7
XX
SQ Sequence 8041 BP; 2241 A; 1807 C; 1838 G; 2126 T; 0 U; 29 Other;

Query Match 100.0%; Score 2091; DB 2; Length 8041;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2091; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGAATAACCACTCTGGCTTCGGTAGTCATTCCTCGTCTCGGATTTTCAGCCAGCAGC 60
DB 3036 ATGCGAATAACCACTCTGGCTTCGGTAGTCATTCCTCGTCTCGGATTTTCAGCCAGCAGC 3095

QY 61 ATAGCTCTCCAGAGGATGTGATTTCTCGCATCCGGCTATGAGAAAGCTGACT 120
DB 3096 ATAGCTCTCCAGAGGATGTGATTTCTCGCATCCGGCTATGAGAAAGCTGACT 3155

QY 121 AACCGACCGCCAGTGTCTGTGATTTAGCCAGAGGAAATTCAGTCCAGCCAGTACCAC 180
DB 3156 AACCGACCGCCAGTGTCTGTGATTTAGCCAGAGGAAATTCAGTCCAGCCAGTACCAC 3215

QY 181 GATCTCGCGAGGCTCTGAGATCAGTAGAGGGTGTGATTTGAAAGTGTGACGGTAA 240
DB 3216 GATCTCGCGAGGCTCTGAGATCAGTAGAGGGTGTGATTTGAAAGTGTGACGGTAA 3275

QY 241 ACCGAGGGCTGGAATACAGATCCAGGAAATGCCAGCAGTTCACAGCTGATCTGATT 300
DB 3276 ACCGAGGGCTGGAATACAGATCCAGGAAATGCCAGCAGTTCACAGCTGATCTGATT 3335

QY 301 GATGGTTCCTCAGGCGGAAAGCAGTACGCTGACTCCCAAGGTTTTCTGCCATGAAT 360
DB 3336 GATGGTTCCTCAGGCGGAAAGCAGTACGCTGACTCCCAAGGTTTTCTGCCATGAAT 3395

QY 361 ACCGGGTTATGCCCCCTCTGGCCGCAATGAGCGTATTCAGGTTATCAGGGGCGCGATG 420
DB 3396 ACCGGGTTATGCCCCCTCTGGCCGCAATGAGCGTATTCAGGTTATCAGGGGCGCGATG 3455

QY 421 TCACACTGTATGGCTCTGATCGGATGGCGGTGTGGTGAATATATTACCAAGAAAT 480
DB 3456 TCACACTGTATGGCTCTGATCGGATGGCGGTGTGGTGAATATATTACCAAGAAAT 3515

QY 481 GCAGACAAATGGCTCTCTCCGTCAATCGAGGGCTGAATCTGCAGGAAGCAAAATGG 540

DB 3516 GCAGACAAATGGCTCTCTCCGTCAATCGAGGGCTGAATCTGCAGGAAGCAAAATGG 3575
QY 541 GGTAAACAGAGCCAGTTAAATTTCTGGAGCAGTGGTCCCTTTGGATGATCTGTGACG 600
DB 3576 GGTAAACAGAGCCAGTTAAATTTCTGGAGCAGTGGTCCCTTTGGATGATCTGTGACG 3635
QY 601 CTGCAAGTACCGGTAGCACACAAACAGCGTCAAGGGTTCATCGGTCACTCATCTGAGCGAT 660
DB 3636 CTGCAAGTACCGGTAGCACACAAACAGCGTCAAGGGTTCATCGGTCACTCATCTGAGCGAT 3695
QY 661 ACAGCAGCAGCGGTATTCCTTATCCACCGAGTCAACAGAAATTAATCTTTGGTGCAGT 720
DB 3696 ACAGCAGCAGCGGTATTCCTTATCCACCGAGTCAACAGAAATTAATCTTTGGTGCAGT 3755
QY 721 CTTGACTGGAAGGCTCGGAGCAGATGTCTCTGGTTTGTATATGATATGATACCAACCGGACG 780
DB 3756 CTTGACTGGAAGGCTCGGAGCAGATGTCTCTGGTTTGTATATGATATGATACCAACCGGACG 3815
QY 781 CGTTATGATAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCCGACCTG 840
DB 3816 CGTTATGATAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCCGACCTG 3875
QY 841 CGCTATGAGCGAACAATAATTCAGCTGGCTATGATCATATCTTTCACCTTCGGAACTGG 900
DB 3876 CGCTATGAGCGAACAATAATTCAGCTGGCTATGATCATATCTTTCACCTTCGGAACTGG 3935
QY 901 AAATCGTATCTGAACTGGAAACGACAGAGAAATAAAGTCTGTAGCTTTGACGAGTGA 960
DB 3936 AAATCGTATCTGAACTGGAAACGACAGAGAAATAAAGTCTGTAGCTTTGACGAGTGA 3995
QY 961 CTGAAGCGCGAACAATAATGGGGCTTGGCGGTGAGCGGGAGCTTAAGGAATCGAACCTT 1020
DB 3996 CTGAAGCGCGAACAATAATGGGGCTTGGCGGTGAGCGGGAGCTTAAGGAATCGAACCTT 4055
QY 1021 ATCTGGAATTCATTACTTCTTACCCCTCTGGGAGAAATCTCATCTGTTTACGTTGGGGGGC 1080
DB 4056 ATCTGGAATTCATTACTTCTTACCCCTCTGGGAGAAATCTCATCTGTTTACGTTGGGGGGC 4115
QY 1081 GAGTTTCAGAGCTCGTCCATGAAAGAGCGAGTTCTTCTGCGACACAGGTGAACTTTC 1140
DB 4116 GAGTTTCAGAGCTCGTCCATGAAAGAGCGAGTTCTTCTGCGACACAGGTGAACTTTC 4175
QY 1141 CGGAGAAAGCTCGTCCGTTATTTCTGAGATCAGTGGCATCTCACGATGACATTCGCG 1200
DB 4176 CGGAGAAAGCTCGTCCGTTATTTCTGAGATCAGTGGCATCTCACGATGACATTCGCG 4235
QY 1201 CTGACTGCGGCGACCGCTATGAAACATCATGAGCAATTCGGGGGACACTTCAGTCCGCGT 1260
DB 4236 CTGACTGCGGCGACCGCTATGAAACATCATGAGCAATTCGGGGGACACTTCAGTCCGCGT 4295
QY 1261 GCATATCTGGTCTGGGATGGGAGATGCCCTGGAGCTGAAAGCGGTGTGACCAACGGGA 1320
DB 4296 GCATATCTGGTCTGGGATGGGAGATGCCCTGGAGCTGAAAGCGGTGTGACCAACGGGA 4355
QY 1321 TATAAGGCAACCCAGAAATGGGCGAGCTACATAAAGGGATTAGTGTGTCTCCGGGACGGGA 1380
DB 4356 TATAAGGCAACCCAGAAATGGGCGAGCTACATAAAGGGATTAGTGTGTCTCCGGGACGGGA 4415
QY 1381 AAAACAAATCTACTTGTGTAAACCCCGACCTGAAGCCGGAAGAGAGCGTCACTTATGAGGCT 1440
DB 4416 AAAACAAATCTACTTGTGTAAACCCCGACCTGAAGCCGGAAGAGAGCGTCACTTATGAGGCT 4475
QY 1441 GGGGTGTATTACGATAAACCCTGGCGGTCTGAAATGCCAATCTCAAGGTTTATTAAGTAC 1500
DB 4476 GGGGTGTATTACGATAAACCCTGGCGGTCTGAAATGCCAATCTCAAGGTTTATTAAGTAC 4535
QY 1501 TTCTCCAAACAGATGTCTCTTATTCATTAATTAACCAATAGTATGATTAACAGC 1560
DB 4536 TTCTCCAAACAGATGTCTCTTATTCATTAATTAACCAATAGTATGATTAACAGC 4595
QY 1561 GGAAGGCCCGGTGACGGTGTGGAATTTGCGGACATTGCGGCTGTGCTGAGAGAT 1620

QY	601	CTGCAGGTACCGCGGTAGCACAACAACAGCGTCAGGGTTTCATCGGTCACTCATCTGACGGAT	660
Db	49391	CTGCAGGTACCGCGGTAGCACAACAACAGCGTCAGGGTTTCATCGGTCACTCATCTGACGGAT	49332
QY	661	ACAGCAGCGACCGGTATTTCTTTATCCACAGGAGTCAAGAAATATATATCTTGGTGCACGT	720
Db	49331	ACAGCAGCGACCGGTATTTCTTTATCCACAGGAGTCAAGAAATATATATCTTGGTGCACGT	49272
QY	721	CTTGACTGGAAGGCGTCGGAGCAGGATGTCTCTGTTTGAATATGATACCAACCCGGCAG	780
Db	49271	CTTGACTGGAAGGCGTCGGAGCAGGATGTCTCTGTTTGAATATGATACCAACCCGGCAG	49212
QY	781	CGTTATCATACCGGATGGCAACTGGGAGGTCACCGGGGGGATATGACCGCGACCCCTG	840
Db	49211	CGTTATCATACCGGATGGCAACTGGGAGGTCACCGGGGGGATATGACCGCGACCCCTG	49152
QY	841	CGCTATCAGCGCAACAAAAATTTGAGTGGCTATGATCATATTTCACTTTCCGAAATCGG	900
Db	49151	CGCTATCAGCGCAACAAAAATTTGAGTGGCTATGATCATATTTCACTTTCCGAAATCGG	49092
QY	901	AAATCTGATCTGTAACCTGGACAGAGACAGAAAAATAAAGGTCGTGAGCTGTACCGAGTCTA	960
Db	49091	AAATCTGATCTGTAACCTGGACAGAGACAGAAAAATAAAGGTCGTGAGCTGTACCGAGTCTA	49032
QY	961	CTGAAGCGCGACAAATGGGGGCTTGGCGGTCAAGCCGGGAGCTTAAAGGAATCGAACCTT	1020
Db	49031	CTGAAGCGCGACAAATGGGGGCTTGGCGGTCAAGCCGGGAGCTTAAAGGAATCGAACCTT	48972
QY	1021	ATCTGTAATTCATCTACTGCTTACCCCTCTGGGAGATCTCATCTGGTTACGGTGGGGGC	1080
Db	48971	ATCTGTAATTCATCTACTGCTTACCCCTCTGGGAGATCTCATCTGGTTACGGTGGGGGC	48912
QY	1081	GAGTTTCAGAGCTCGTCCATGAAAGACGGAGTTGTCTTTGCCAGACACAGGTGAAACTTTC	1140
Db	48911	GAGTTTCAGAGCTCGTCCATGAAAGACGGAGTTGTCTTTGCCAGACACAGGTGAAACTTTC	48852
QY	1141	CGGCAGAAAAGCTGCTCGGTATTTGCTGAGGATGAGTGGCATCTCACCGATGCACCTTCG	1200
Db	48851	CGGCAGAAAAGCTGCTCGGTATTTGCTGAGGATGAGTGGCATCTCACCGATGCACCTTCG	48792
QY	1201	CTGACTCGCGGCAGCCGCTATGAAATCATGAGCAATTCGGGGACACTTCAGTCCGCGT	1260
Db	48791	CTGACTCGCGGCAGCCGCTATGAAATCATGAGCAATTCGGGGACACTTCAGTCCGCGT	48732
QY	1261	GCATATCTGCTCTGGGATGTGGCAGATGCTCTGGAACCTGAAAGGCGGTGTGACACGGGA	1320
Db	48731	GCATATCTGCTCTGGGATGTGGCAGATGCTCTGGAACCTGAAAGGCGGTGTGACACGGGA	48672
QY	1321	TATAAGGCACCCAGAAATGGGGCAGCTACATAAAGGATTAGTGGTGTGTCTCGGGCAGGGA	1380
Db	48671	TATAAGGCACCCAGAAATGGGGCAGCTACATAAAGGATTAGTGGTGTGTCTCGGGCAGGGA	48612
QY	1381	AAAAAATACTACTTGGTAAACCCCGACCTTGAAGCCGGGAAGAGCGTCAGATTAGAGCT	1440
Db	48611	AAAAAATACTACTTGGTAAACCCCGACCTTGAAGCCGGGAAGAGCGTCAGATTAGAGCT	48552
QY	1441	GGGTGTATTACGATAACCCCGCGGTCTGGAATGCCAATGTCAAGGTTTTTATGACTGAC	1500
Db	48551	GGGTGTATTACGATAACCCCGCGGTCTGGAATGCCAATGTCAAGGTTTTTATGACTGAC	48492
QY	1501	TTTCTCAACAAGATTTGCTCTTTATTCCTAATAATGATTAACACAATAAGTATGTATAACAGC	1560
Db	48491	TTTCTCAACAAGATTTGCTCTTTATTCCTAATAATGATTAACACAATAAGTATGTATAACAGC	48432
QY	1561	GGAAAGCCCGGTTGCACGGTGTGGAATTTGCCGGCACATTCGCCCTCTGGTTCAGAGGAT	1620
Db	48431	GGAAAGCCCGGTTGCACGGTGTGGAATTTGCCGGCACATTCGCCCTCTGGTTCAGAGGAT	48372
QY	1621	GTCAAGCTGTCACTGAATTACCTCGAACCCGAAGTGAAACACAGCTGATGGTGTATAACAAA	1680
Db	48371	GTCAAGCTGTCACTGAATTACACCTCGAACCCGAAGTGAAACACAGCTGATGGTGTATAACAAA	48312

QY	1891	GGTGGCGCGTGGATTATACCCCTGAAACACATGGTGAATGCCAAACTGAACTGGCAGATC	1740
DB	48311	GGTGGCGCGTGGATTATACCCCTGAAACACATGGTGAATGCCAAACTGAACTGGCAGATC	48352
QY	1741	ACCGAAGAGTGGCATCATGCTGGGTGGCCGGTTATCGCGGGGAAACACACAGTTTCACC	1800
DB	48351	ACCGAAGAGTGGCATCATGCTGGGTGGCCGGTTATCGCGGGGAAACACACAGTTTCACC	48192
QY	1801	CAGAAATTATTCGTACATGAGCGGTGTACAGAGAAAGTGTATGATGAGAAAGGAGATAC	1860
DB	48191	CAGAAATTATTCGTACATGAGCGGTGTACAGAGAAAGTGTATGATGAGAAAGGAGATAC	48132
QY	1861	CTGAAAGCCTGGACGGTGGTGGAGTGCAGGTCTGTCTGTGGAAGATGACGGATGCCCTGACG	1920
DB	48131	CTGAAAGCCTGGACGGTGGTGGAGTGCAGGTCTGTCTGTGGAAGATGACGGATGCCCTGACG	48072
QY	1921	CTGAATGCTCGGTGAATAACCTGTCTCAACAAAGGATTACAGTGCAGTGCAGCTGTACAGT	1980
DB	48071	CTGAATGCTCGGTGAATAACCTGTCTCAACAAAGGATTACAGTGCAGTGCAGCTGTACAGT	48012
QY	1981	GCCGGTGAAGTACGCTGTATGCGGTGATTACTTCCAGACGGGATCATCACACAGGA	2040
DB	48011	GCCGGTGAAGTACGCTGTATGCGGTGATTACTTCCAGACGGGATCATCACACAGGA	47952
QY	2041	TATGTGATACCTGAGCGAAATTACTGATGTCGCTCAACTATCAGTTCTGA	2091
DB	47951	TATGTGATACCTGAGCGAAATTACTGATGTCGCTCAACTATCAGTTCTGA	47501
RESULT 3			
ADC00087/c			
ID	ADC00087	standard; DNA; 86248 BP.	
AC	ADC00087;		
XX			
DT	04-DEC-2003	(first entry)	
DE	Enterohaemorrhagic E. coli O157:H7-specific nucleic acid	SEQ ID NO: 132.	
KW	ds; gene; enterohaemorrhagic; anti-bacterial.		
XX			
OS	Escherichia coli; O157:H7.		
XX			
PN	JP2002355074-A.		
XX			
PD	10-DEC-2002.		
XX			
PF	24-JAN-2002; 2002JP-00015959.		
XX			
PR	24-JAN-2001; 2001JP-00112010.		
XX			
PA	(UUTS-) UNIV TSUKUBA.		
XX			
DR	WPI; 2003-451640/43.		
XX			
PT	Enterohaemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule		
PT	and a polypeptide and its use, a polypeptide, a vector and a host cell.		
XX			
PS	Claim 2; SEQ ID NO 132; 2067pp; Japanese.		
XX			
CC	The invention relates to a novel enterohaemorrhagic Escherichia coli		
CC	O157:H7-specific nucleic acid molecule. A polynucleotide of the invention		
CC	has anti-bacterial activity. The polypeptide can be used in detection		
CC	and/or treatment of O157:H7 infection. The nucleotide sequence of the		
CC	genome of Enterohaemorrhagic E. coli O157:H7 was determined. The present		
CC	sequence represents an E. coli O157:H7-specific nucleic acid of the		
CC	invention.		
XX			
SQ	Sequence 86248 BP; 22338 A; 20003 C; 21303 G; 22604 T; 0 U; 0 Other;		
Query Match			
Best Local Similarity 99.8%; Score 2087.8; DB 9; Length 86248;			
Matches 2089; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			

AC ABS78942;
 XX 17-DEC-2002 (first entry)
 XX DE E. coli CFT073 genomic sequence #109.
 XX Pathogenic; Escherichia coli CFT073 infection; livestock; pyelonephritis;
 KW urinary tract infection; open reading frame; ORF; uropathogenic;
 XX antibacterial; atrophic; nephrotropic; gene; ds.
 XX Escherichia coli.
 XX WO200259320-A2.
 XX PN 01-AUG-2002.
 XX PD 19-OCT-2001; 2001WO-US046833.
 XX PF 19-OCT-2000; 2000US-0242412P.
 XX PR (WISC) WISCONSIN ALUMNI RES FOUND.
 XX PA Blattner FR, Welch RA, Burland VD;
 XX PI WPI; 2002-691532/74.
 XX DR New DNA sequences of the pathogenic Escherichia coli CFT073 strain,
 XX PT useful for preventing or treating E. coli CFT073 infection in humans or
 XX PT livestock.
 XX PS Claim 1; Page 323-366; 765pp; English.
 XX The present invention relates to polynucleotide sequences from the genome
 CC of the pathogenic Escherichia coli strain CFT073. Almost all the
 CC sequences present in E. coli CFT073 are absent in the previously
 CC sequenced laboratory strain K-12. The polynucleotide sequences of the
 CC invention are useful for preventing, diagnosing or treating E. coli
 CC CFT073 infection in humans or livestock. The polynucleotide sequences are
 CC useful for preventing urinary tract infections and pyelonephritis.
 CC Likewise, the polypeptides encoded by the different open reading frames
 CC (ORF1-5) are useful for generating a vaccine against uropathogenic E.
 CC coli strains. ABS78834-ABS79085 represent genomic sequences from E. coli
 CC strain CFT073
 XX
 SQ Sequence 76804 BP; 20476 A; 18524 C; 17417 G; 20360 T; 0 U; 27 Other;
 Query Match 99.5%; Score 2075.8; DB 6; Length 76804;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2084; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 1 ATGGGAATACCACTCTGGCTTCGGTAGTCAATCCCTGCTCGGATTTTCAGCCAGCAGC 60
 DB 44573 ATGGGAATACCACTCTGGCTTCGGTAGTCAATCCCTGCTCGGATTTTCAGCCAGCAGC 44514
 QY 61 ATAGCTGCTGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
 DB 44513 ATAGCTGCTGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 44454
 QY 121 AACGACGCGCAGTGTTCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
 DB 44453 AACGACGCGCAGTGTTCGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 44394
 QY 181 GATCTGCGGAGGCTCTGAGATCAGTAGAGGCTGATGATGATGATGATGATGATGATGATGAT 240
 DB 44393 GATCTGCGGAGGCTCTGAGATCAGTAGAGGCTGATGATGATGATGATGATGATGATGAT 44334
 QY 241 ACCGAGGCTGGAATCAGATCCGAGGAATGCCAGCAGGATGATGATGATGATGATGATGAT 300
 DB 44333 ACCGAGGCTGGAATCAGATCCGAGGAATGCCAGCAGGATGATGATGATGATGATGATGAT 44274
 QY 301 GATGGTGTTCGTGAGGCGGAAGCAGTACGATGATGATGATGATGATGATGATGATGATGAT 360
 DB 44273 GATGGTGTTCGTGAGGCGGAAGCAGTACGATGATGATGATGATGATGATGATGATGAT 44214

QY 361 ACCGGTTTCATGCCCCCTCTGCGCCCATTTGAGCGTATTGAGGTTATCAGGGGCGCGATG 420
 DB 44213 ACCGGTTTCATGCCCCCTCTGCGCCCATTTGAGCGTATTGAGGTTATCAGGGGCGCGATG 44154
 QY 421 TCCACACTGTATGCTCTGATGCGATGGCGGTCGTGTTGTAATATCATTTACAGAAAGAT 480
 DB 44153 TCCACACTGTATGCTCTGATGCGATGGCGGTCGTGTTGTAATATCATTTACAGAAAGAT 44094
 QY 481 GCAGACAAATGCTCTCTCCGTCAATCCAGGCTGATCTGCAGGAAAGCAAAATGG 540
 DB 44093 GCAGACAAATGCTCTCTCCGTCAATCCAGGCTGATCTGCAGGAAAGCAAAATGG 44034
 QY 541 GGTAAACAGCAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCTTGTGATGATTTCTGTGATG 600
 DB 44033 GGTAAACAGCAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCTTGTGATGATTTCTGTGATG 43974
 QY 601 CTGACAGTACCGGTAGCACAACAACAGGCTCAGGTTTCATCGGTCAATCATCTGAGGAT 660
 DB 43973 CTGACAGTACCGGTAGCACAACAACAGGCTCAGGTTTCATCGGTCAATCATCTGAGGAT 43914
 QY 661 ACAGCAGCAGCGTATTCCTTATCCCAACGAGTCAAGAAATATAATCTTGGTGACAGT 720
 DB 43913 ACAGCAGCAGCGTATTCCTTATCCCAACGAGTCAAGAAATATAATCTTGGTGACAGT 43854
 QY 721 CTGACAGTACCGGTAGCACAACAACAGGCTCAGGTTTCATCGGTCAATCATCTGAGGAT 780
 DB 43853 CTGACAGTACCGGTAGCACAACAACAGGCTCAGGTTTCATCGGTCAATCATCTGAGGAT 43794
 QY 781 CGTTATGATTAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCTG 840
 DB 43793 CGTTATGATTAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCTG 43734
 QY 841 CGTTATGATTAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCTG 900
 DB 43733 CGTTATGATTAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCTG 43674
 QY 901 AAATCGTATCTGAATGGAACGACAGCAAGAAATGAAGTCTGAGCTTGTACGAGTGA 960
 DB 43673 AAATCGTATCTGAATGGAACGACAGCAAGAAATGAAGTCTGAGCTTGTACGAGTGA 43614
 QY 961 CTGAAGCGCGCAAAATGGGGCTTTCGGGTGACGGGGGAGCTTAAGGAATCGAACCTT 1020
 DB 43613 CTGAAGCGCGCAAAATGGGGCTTTCGGGTGACGGGGGAGCTTAAGGAATCGAACCTT 43554
 QY 1021 ATCCGTAATTCATTTACTGCTTACCCCTCTGGGAGAAATCTCATCTGTTACGTTGGGGGC 1080
 DB 43553 ATCCGTAATTCATTTACTGCTTACCCCTCTGGGAGAAATCTCATCTGTTACGTTGGGGGC 43494
 QY 1081 GAGTTTCAGAGCTCGTCCATGAAAGAGGAGTTCCTTCCACACAGAGTGAACCTTTC 1140
 DB 43493 GAGTTTCAGAGCTCGTCCATGAAAGAGGAGTTCCTTCCACACAGAGTGAACCTTTC 43434
 QY 1141 CGGAGAAAGCTGGTGGTATTTGCTGAGGATGAGTGGATCTCAGGATGCACTTGGC 1200
 DB 43433 CGGAGAAAGCTGGTGGTATTTGCTGAGGATGAGTGGATCTCAGGATGCACTTGGC 43374
 QY 1201 CTGACTGCGGCGCAGCCGTATGAACATCATGAGCAATTCGGGGGACACTTCAGTCCGCT 1260
 DB 43373 CTGACTGCGGCGCAGCCGTATGAACATCATGAGCAATTCGGGGGACACTTCAGTCCGCT 43314
 QY 1261 GCATATCTGGTCTGGGATGTGGCAGATGCTTGGACGCTGAAAGCGGTGTGACACAGGGA 1320
 DB 43313 GCATATCTGGTCTGGGATGTGGCAGATGCTTGGACGCTGAAAGCGGTGTGACACAGGGA 43254
 QY 1321 TATAGGCAACCCAGAAATGGGCGACATACATAAGGATAGTGTGTCCGGGACAGGA 1380
 DB 43253 TATAGGCAACCCAGAAATGGGCGACATACATAAGGATAGTGTGTCCGGGACAGGA 43194
 QY 1381 AAAACAAATCTACTTGTGTAAACCCGACCTGAAGCCGGAAGAGCGGTCACTGATGAGCT 1440
 DB 43193 AAAACAAATCTACTTGTGTAAACCCGACCTGAAGCCGGAAGAGCGGTCACTGATGAGCT 43134

RESULT 6
ACF65383_0
WP Sequence split into 5 fragments LOCUS ACF65383 Accession ACF65383
WP Fragment Name Begin End
ACF65383_0 1 110000
WP ACF65383_1 100001 210000
WP ACF65383_2 200001 310000
WP ACF65383_3 300001 410000
WP ACF65383_4 400001 460203
ID ACF65383 standard; DNA; 460203 BP.
XX
AC ACF65383;
XX
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens nucleotide sequence #36.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough; gene; ds.
XX
OS Photorhabdus luminescens.
XX
XX WO200294867-A2.
XX
PN 28-NOV-2002.
XX
PP 07-FEB-2002; 2002WO-IB003040.
XX
XX 07-FEB-2001; 2001FR-00001659.
XX
XX (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX
XX Duchaud E, Tacurit S, Glaser P, Frangeul L, Kunst F, Danchin A;
XX Buchrieser C;
XX
XX WPI; 2003-148459/14.
XX
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
XX Claim 1; SEQ ID NO 36; 1205pp; French.
XX
XX The invention relates to the isolation of genes and their encoded
XX proteins from Photorhabdus luminescens. The isolated sequences are
XX sources of probes and primers for detecting the genome of P. luminescens
XX and related species; to study polymorphisms; for gene analysis and for
XX detection/amplification of the genes. Antibodies (Ab) raised against the
XX polypeptides encoded by the genes are used for detection/identification
XX of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX carry a gene-containing vector are used to select compounds that
XX modulate, regulate, induce or inhibit expression of the genes in plants,
XX animals or microorganisms other than P. luminescens and are able to alter
XX response or sensitivity to toxins and antibiotics produced by P.
XX luminescens. Cells transformed to express the genes are useful for
XX recombinant production of the proteins, particularly toxins and
XX antibacterials useful as insecticides, bactericides and fungicides. The
XX genes, proteins, vectors containing the genes and Ab are also useful
XX therapeutically (to treat microbial infection by bacteria or fungi that
XX are sensitive to P. luminescens-encoded toxins or antibiotics) and as
XX biopesticides. Other uses of the genes and the proteins are as virulence
XX factors and for identifying targets of human diseases for which P.
XX luminescens is a model (particularly plague and whooping cough). This
XX sequence represents one of the isolated P. luminescens genes
XX
SQ Sequence 460203 BP; 128022 A; 88971 C; 107007 G; 136199 T; 0 U; 4 Other;
Query Match 7.1%; Score 149.4; DB 7; Length 110000;
Best Local Similarity 57.0%; Pred. No. 1.2e-35;

Matches 298; Conservative 0; Mismatches 216; Indels 9; Gaps 1;
QY 92 CGGCATCCGGCTATGAGAAAAAGCTGACTAAACGAGCGCGCAGTGTCTTCTGTATTAGCC 151
DB 58493 CGGCAGCAGGTTTTCACAGAGAATTGAAGATGCGCCAGCGCTATTCTGTGTGAGCC 58552
QY 152 AGGAGGAATTTCAGTCCAGCCAGTACACAGATCTGGCGAGGCTCTGAGATCAGTAGAGG 211
DB 58553 GTGACCAATTGGAACTAAGCGTTATCGTAGTGTACTGATGCTGAAAGATGTACCGG 58612
QY 212 GTGTGGATGTTGAAAGTGTACGGGTAAACCGGAGGCTGGAATACAGCATCCGAGGAA 271
DB 58613 GTGTGTTTACAGGTGTTGCTAGCA-----GCAGTGATATCAGTATTCGCGGTA 58663
QY 272 TCCAGCCAGTTACAGCGTGATCTACTGATTGATGTTGCTCAGCGGGAAGCAGTAGACG 331
DB 58664 TGGCGGCTCAATATACAAATGATTTGGTTGATGTTAAGCGTGTGATACCGGAGGTACTC 58723
QY 332 TGACTCCCAACGGTTTTTCTGCCATGAATACCGGGTTTCATGCCCTCTTGGCCGCCATTG 391
DB 58724 GCCCAACCGTGATAATTCGGGGATTGAACAGGGGTGGTTGCCCGCATTCGCCGATAG 58783
QY 392 AGCGTATTGAGGTTATCAGGGGGCGGATGTCCACACTGTATGGCTCTGATCGCATGGGCG 451
DB 58784 AACGTATCGAGGTTGTTGCTGCTCTATGCTCTCTTTATGTTCTGTATCGCATGGGCG 58843
QY 452 GTGTGGTGAATATCATTTACCAAGAAAGATGCAGACAAATGGCTCTCTTCCTCAATCAG 511
DB 58844 GGTGATTTAATCATCATTTACCGGTAAAGCAGAGAAGAGTGAACACAGCTTCGCCGCTG 58903
QY 512 GGCTGAATCTGCAGAAAGCAAAATGGGGTAAACAGCAGCCAGTTTAATTTCTGGAGCA 571
DB 58904 ATGGCACTTTCAGGAGGTTTCTAAATCAGGTAATAGTATCATCAGAGTAGTGTGTATATT 58963
QY 572 GTGTGTCCTCTGTGGATGATTCTCTCAGCCCTGAGGTACGCGG 614
DB 58964 CAGGTCCAATTATTGATGGATTATTAGGGCTGAAAGTAGGCGG 59006

RESULT 7

ACF67367_01
Continuation (2 of 57) of ACF67367 from base 100001 (Photorhabdus luminescens nucleotide
WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367
WP Fragment Name Begin End
ACF67367_00 1 110000
WP ACF67367_01 100001 210000
WP ACF67367_02 200001 310000
WP ACF67367_03 300001 410000
WP ACF67367_04 400001 510000
WP ACF67367_05 500001 610000
WP ACF67367_06 600001 710000
WP ACF67367_07 700001 810000
WP ACF67367_08 800001 910000
WP ACF67367_09 900001 1010000
WP ACF67367_10 1000001 1110000
WP ACF67367_11 1100001 1210000
WP ACF67367_12 1200001 1310000
WP ACF67367_13 1300001 1410000
WP ACF67367_14 1400001 1510000
WP ACF67367_15 1500001 1610000
WP ACF67367_16 1600001 1710000
WP ACF67367_17 1700001 1810000
WP ACF67367_18 1800001 1910000
WP ACF67367_19 1900001 2010000
WP ACF67367_20 2000001 2110000
WP ACF67367_21 2100001 2210000
WP ACF67367_22 2200001 2310000
WP ACF67367_23 2300001 2410000
WP ACF67367_24 2400001 2510000
WP ACF67367_25 2500001 2610000
WP ACF67367_26 2600001 2710000
WP ACF67367_27 2700001 2810000
WP ACF67367_28 2800001 2910000


```

QY 422 CCACACTGTATGGCTCTGATCGATGGCGGTGTGGTGAATATCATTTACCAGAAAGAATG 481
Db 9757 CGTCGCTGTACGGTTCGATGCGCTCGCGGTGTAGTATATCATCACCAAAAATCG 9698
QY 482 CAGACAAATGGCTCTCTCCGTCGAATGACAGGGTGAATCTGACGAGAAAGCAACAATGGG 541
Db 9697 GTACAGAAATGTCGGGTACCGTTACCGTTCGATACCAACCATTCAGGAACATCGCGATCGCG 9638
QY 542 GTACACACAGCCGCTTTAAATTTCTGGAGCAGTGGTCCCTTGTGGATGATCTGTGACGC 601
Db 9637 GTGACACTTAACCGTCACTTTTACCAGTGAACCAATTAATGATGGTGTCTGGAA 9578
QY 602 TGACAGTACGCGGTAGCACACAACAGCGTCA 632
Db 9577 TGAAGCTTACGCGCGCTCGCAAAACGTGA 9547

RESULT 9
ABA89269
ID ABA89269 standard; DNA; 2046 BP.
XX AC ABA89269;
XX DE
XX DT 11-FEB-2002 (first entry)
XX ES Escherichia coli polynucleotide SEQ ID NO 1076.
XX KW Escherichia coli; B2/D-A; antiinflammatory; antibacterial;
XX KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
XX KW systemic infection; non-diarrhoeal infection; septicemia;
XX KW pyelonephritis; antibiotic resistance; ds.
XX OS Escherichia coli.
XX PN WO200166572-A2.
XX PD 13-SEP-2001.
XX PF 12-MAR-2001; 2001WO-BP003445.
XX PR 10-MAR-2000; 2000FR-00003145.
XX PR 02-FEB-2001; 2001FR-00001449.
XX PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
XX WPI; 2001-550253/61.
XX PT A library of DNA fragments of Escherichia coli strains for the phylogenetic
XX FT determination of a given strain comprises polynucleotides of nature B2/D+
XX A-.
XX PS Example 6; Fig 6; 646pp; English.
XX CC The invention relates to a library of DNA fragments of Escherichia coli
XX CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533) and
XX CC encoded proteins (ABE52459-ABE52919 and ABE52954-ABE53094) of nature
XX CC B2/D+A-. The polynucleotides have potential antiinflammatory,
XX CC antibacterial and immunosuppressive activity as part of pharmaceutical
XX CC compositions used to treat, palliate or prevent extra-intestinal E. coli
XX CC infections. The polypeptides are useful for determining the phylogenetic
XX CC group of a given E. coli strain. These polypeptides can detect and treat
XX CC an undesired development of E. coli, particularly an extra-intestinal
XX CC infection that include systemic and non-diarrhoeal infections such as
XX CC septicemia, pyelonephritis and meningitis this is particularly
XX CC advantageous as bacterial resistance is increasing with the more frequent
XX CC use of broad spectrum antibiotics
SQ Sequence 2046 BP; 645 A; 396 C; 484 G; 521 T; 0 U; 0 Other;
Query Match 6.0%; Score 125.6; DB 4; Length 2046;

```

```

Best Local Similarity 46.3%; Pred. No. 5.4e-29;
Matches 684; Conservative 0; Mismatches 759; Indels 33; Gaps 7;
QY 72 AGAGATGTGATGATTGTCGCGCATCGGCTATGAGAAAAGCTGACTAACGAGCGCG 131
Db 87 AGAAGATACGGTTGTTGTTACTGCTCGGGTTCACCTCAGCAGCTCAGAATGCCCGGC 146
QY 132 CAGTGTCTCTGTGATTACCCAGGAGGAATGCACTCCAGCAGTACCAAGATCTGGCGGA 191
Db 147 CAGTGTCTCAGTCATTACTTCAGAAACAATGCAAAAAAACCAGGTTTCAGATCTGGTCGA 205
QY 192 GGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAGTGGTACGGGTAAACCCGAGGGCT 251
Db 207 TG-----CAGTAAAGATGTTGAAGGATTAGTATCCTGCTGGGAATGAAAAACC 257
QY 252 GGAATCAGCATCCGAGGAATGCCAGCCAGTTACACGCTGATACCTGATGATGTTGTCG 311
Db 258 GGAATCAGTATACGTGCTAAGTGGCGATTACACGCTGATTCTGCTCGATGGACGACG 317
QY 312 TCAGGCGGAAGCAGTAGTACGCTCCCAACGGTTTTCTGCCATGAATACCGGTTTCA 371
Db 318 TCAGAGCG--GTCCGGGAATCCAGCAACGCGGCGGGTTTTGAAGCGGATTTAT 374
QY 372 GCCCCTCTGCCCGCCATTGAGCGTATTGAGTTTATCAGGGGGCGGATGCCACACTGTA 431
Db 375 CCTCTCTGGAAGCAATTGAACGCAATTGAAGTATCGGTGGCCCTATGTTCTCCCTGTA 434
QY 432 TGGCTCTGATCGGATGGCGGTGTGGTGAATATATCATTACAGAAAGAATCGACAAATG 491
Db 435 TGGTTCTGATGCCATCGAGGGGTCAATTAATATATCAACCAACCACTTAATAACCAAA 494
QY 492 GCTCTCTCCGTCATGCA---GGGCTGAATCTCAGGAAAGCAACAAATGGGGTAAACAG 548
Db 495 ATGGGATGGCGTACTTGGACTTTGGGGGATTTTCAGGAACATGGGAAATTTGGTAACTC 554
QY 549 CAGCCAGTTTAAATTTCTGGAGCAGTGTGCCCTTGTGGATGATTCTGTACGCTGACGTT 608
Db 555 AACCAAAATGACTTCTATCTGTCCAGGCCCATTTGATTAAAGGATAAACTTGTCTTCAGCT 614
QY 609 ACGGGTAGCACACAACAGGCTCAGGGTTTCATCGTCCATCATCTGAGAGCATACAGCAGG 668
Db 615 ATATGGAGGAATGAACCTATCCAAAG-----AAGATAGTATCTCTCAGGA-AACCGCG 668
QY 669 CACGCTATTCTCTTATCCACGGAGTCAAGAAATTAATCTTGGTGCACGCTCTTGAAGT 728
Db 669 AAAGATAATAAGAAATATAACGGCAACGCTCCAGTTTACTCCGACTGAAAGCCAGAAAGTT 728
QY 729 GAAGCGTCCGAGCAGGATGTGCTGTGGTTGATATGATACCAACCCGAGCGGTTATGA 788
Db 729 TGTTTTGAATATATGAAAAAATAACAGGTGCATACATTAAACACCTGGTGAAGTCTCTCGA 788
QY 789 TAACCGGATGGCAACTGGGGAGTCTGACGGGGGATATGACCGGACCTTCGCTATGA 848
Db 789 TGCCTGGACTATGCGGGAATCTTAAACACCAACAGTAAAGAGAAAGCATATTC 848
QY 849 GCGAAACAAAATTTCACTGGCTATGATCATCTTCACTTCCGAAACATGGAATCGTA 908
Db 849 ACGTAGTCACTCGGTAGCAGCAT-----GGAATGCCCGCGGCGAAATACCTGCACTCTCGA 902
QY 909 TCTGACTGGAACGAGACAGAAAAATAAAGTTCGTGAGCTTGTACGCGAGTGTACTGAAAGCG 968
Db 903 AATTGCTGTTTATCAGGAGAAAGTTTATTCGTAGTTAATCAGTAAAAAGATAATA 962
QY 969 CGACAAATGGGGCTTCCCGGTACGCGCGGAGCTTAAGAAATCGAAACCTTATCTGAA 1028
Db 963 TAATCATTTGGGATCTTAATACGAGTCAAGAAAAACCGAAATAACCAACAATCATAGA 1022
QY 1029 TCAATTACTGTTACCCCTCTGGGAGATCTCATCTGTTACGTTGGGGGGCGAGTTTCA 1088
Db 1023 TGAAGAGTGACGGCATTTCTGCCGGAATTTACTGACCATCGGAGTCAATTTTCAGCA 1082
QY 1089 GAGCTCGTCCATGAAGACGCGAGTTGCTTTCGCCAGCACAGGTGAAA-----CTTTCCG 1142

```


Db 1083 TGACAGCTCCGTGATGACTACGCCCGGTAAAAAAGCAGCAACACAGCTCTGTTTC 1142
 QY 1143 GCAGAAAAGCTGGTGGTATTTCTGAGATGATGGCATCTCAGGATGCACTTGGCT 1202
 Db 1143 AATTAACACAGAGCTGTTTATAGAAATGATATGAGCAACGATTTCTCGCCT 1202
 QY 1203 GACTGGGGGAGCGGTATGAACATCATGAGCAATTCGGGGGACACTTCAGTCCGGTGC 1262
 Db 1203 GACTGGAGGAGCTGCTCGATAATCATGAAATCTATGACAGTTACTGGAATCCAAGATT 1262
 QY 1263 ATATCTGGTCTGGATGTGACAGATCCTGGACGCTGAAAGCGGTGTCACACGGGATA 1322
 Db 1263 GTAGCTGTTTAACTGACCGATATCTCACATCAAGAGGGGATGCAAGACATT 1322
 QY 1323 TAAGGACCCAGAAATGGGAGCTACATAAAGGATTAAGTGTCTCCGGGACAGGAAA 1382
 Db 1323 TCGGGCTCTTCAATTCGTGAGGTAGTCTGATTTGAAACACTGACGAGGGTGGTGC 1382
 QY 1383 AACAAATCTACTTGTAAACCCGACCTGAACCGGAGAGCGGTGATGAGGCTGG 1442
 Db 1383 CTCTATTATGATGGAACACAGGACCTGAAACCGGAGACCGATGTAACCGAAGATCGG 1442
 QY 1443 GGTGTATTACGATAACCCCGCGCTCTGAATGCCAATGTCAACAGGTTTATGACTGACTT 1502
 Db 1443 TATTATTATGATGATAGTGTGTTTTCGGGAGCGGACGCTGTTTAACTGATTT 1502
 QY 1503 CTCACACAGATGCTCTCTATTTCATTAATGATAA 1538
 Db 1503 TAAATAAAGTTGACCAAGTTACGATATAGTACAAA 1538

RESULT 10

AB89255
 ID AB89255 standard; DNA; 12943 BP.
 XX
 AC AB89255;
 XX
 DT 11-FEB-2002 (first entry)
 XX
 DE Escherichia coli polynucleotide SEQ ID NO 1049.
 XX
 KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
 KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
 KW systemic infection; non-diarrhoeal infection; septicaemia;
 KW pyelonephritis; antibiotic resistance; ds.
 XX
 OS Escherichia coli.
 XX
 PN WO200166572-A2.
 XX
 PD 13-SEP-2001.
 XX
 PF 12-MAR-2001; 2001WO-EF003445.
 XX
 PR 10-MAR-2000; 2000PF-00003145.
 XX
 PR 02-FEB-2001; 2001PF-00001449.
 XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
 XX
 DR WPI; 2001-550253/61.
 XX

A library of DNA fragments of Escherichia coli strains for the phylogenetic determination of a given strain comprises polynucleotides of nature B2/D+ A-.

Example 6; Fig 6; 646pp; English.

The invention relates to a library of DNA fragments of Escherichia coli strains comprising polynucleotides (ABA89577-ABA88729 and ABA89533) and encoded proteins (AB52459-AB52919 and AB52954-AB53094) of nature B2/D+A-.

The polynucleotides have potential antiinflammatory,

CC antibacterial and immunosuppressive activity as part of pharmaceutical compositions used to treat, palliate or prevent extra-intestinal E. coli infections. The polypeptides are useful for determining the phylogenetic group of a given E. coli strain. These polypeptides can detect and treat an undesired development of E. coli, particularly an extra-intestinal infection that include systemic and non-diarrhoeal infections such as septicemia, pyelonephritis and meningitis this is particularly advantageous as bacterial resistance is increasing with the more frequent use of broad spectrum antibiotics

XX
 SQ Sequence 12943 BP; 3358 A; 2856 C; 3166 G; 3563 T; 0 U; 0 Other;

Query Match 6.0%; Score 125.6; DB 4; Length 12943;
 Best Local Similarity 46.3%; Pred. No. 1.6e-28;
 Matches 684; Conservative 0; Mismatches 75; Indels 33; Gaps 7;
 QY 72 AGAGATGTGATGTTGTCGGCATCCGGTATGAGAAAAGCTGACTAACGAGCGC 131
 Db 10799 AGAAGATACGCTTGTGTACTGCTCGGTTCTCAGCAGCTCAGAATGCCCGC 10858
 QY 132 CAGTGTTCGTGATTTAGCCAGAGGAATTGCACTCCAGCCAGTACCAGTCTGGCGGA 191
 Db 10859 CAGTGTCTCAGTCACTTACTTCAGAACAACTGCAAAAAAACCCTTCAGATCTGGTCGA 10918
 QY 192 GGCCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGGTACGGGTAAAAACCGAGGCT 251
 Db 10919 TG-----CAGTAAAGATGTTGAAGGATTTAGTATCACTGGTGGATGAAAACC 10969
 QY 252 GGAATCAGCATCCGAGGAATGCCAGCAGTTACGCTGATGATGATGATGATGATG 311
 Db 10970 GGATATCAGTATAGTGTGCTAAGTGGCGATTACACGCTGATTCGTGTCGAGCAGC 11029
 QY 312 TCAGGGCGGAGCAGTACGCTGCTCCCAAGGTTTTTCTGCCATGATACCGGGTTTCA 371
 Db 11030 TCAGAGCG---GTCGGGAATCCAGACCAACCGCAGCGCGGTTTTGAAGCGGATTTAT 11086
 QY 372 GCGCCCTCTGCGCCCATTTGAGCGTATTGAGGTTTACAGGGGCGCGATGTCCACACTGTA 431
 Db 11087 CCTCTGTGGAAGCAATTGAACGATTTGAAGTATCGTGCCTATGTTCTCCCTGTA 11146
 QY 432 TGGCTCTGATCCGATGGCGGTGTGGTGAATATCATTTACAGAAAGATGAGCAAAATG 491
 Db 11147 TGGTCTGATCCATCGGAGGGGTCATTAATATATAACCAACAGTAAATACCAAAAC 11206
 QY 492 GCTCTCTCCGTCAATGCA---GGGCTGAATCTGAGAAAGCAACAAATGGGTAAACAG 548
 Db 11207 ATGGGATGGCTACTTTGGACTTTGGGGGATTTTTCAGGAACATGGAAATTTGGTAATC 11266
 QY 549 CAGCCAGTTTAAATTTTGGAGCAGTGTCCCTTGTGGATGATTTCTGTGACCTTCAGGT 608
 Db 11267 AACCAAAATGACTTCTATCTGTCCAGGCCCATTTGATTAAGGATAAACTTGTCTTCAGCT 11326
 QY 609 ACGCGTAGCACACACAGCGTCCAGGTTTCATCGTCCATCCTGAGCGATACAGCAGG 668
 Db 11327 ATATGGAGGATGAATCTATCCAAAG-----AAGATAGTATCTCTCAGGGA-ACACCGGC 11380
 QY 669 CACGCTATTCCTTATCCACGGAGTACAGAAATTAATCTTGGTGCAGCTTTGACTG 728
 Db 11381 AAAAGATAATAAGAAATATAACGCAACGCTCCAGTTTACTCCGACTGAAGCCAGAAATT 11440
 QY 729 GAAGCGTCGAGCAGGATGTCTGTGTTTGTATGATGATACCAACCGCGAGCGGAGCTTATGA 788
 Db 11441 TGTTTTGAATATGAAAAAATAAACAGGTGATACATTAAACCTGGTGTGCTCTCGA 11500
 QY 789 TAAACGGGATGGCAACTGGGGAGTCTGAACGGGGGATATGACCGGACCTTGGCCTATGA 848
 Db 11501 TGCCTGGACTATCGGGGAATCTTAAACACCAACAGTAAGAGAAAGCAATATTC 11560
 QY 849 GCGAAACAAAATTCAGCTGGCTATGATCATCTTTTTCACCTTCGGACATGGAATCGTA 908
 Db 11561 ACGTAGTCACTGGGTAGCAGCAT-----GGAATGCCAGGCGGAAATCTGTCATCTCGA 11614
 QY 909 TCTGAACGAGCAGACAGAAATAAAGTGTGAGCTTTGTACGAGTGTACTGAAGCG 968

Db 11615 AATTGCTGTTTATCAGGAGAAAGTTATTCGTGAGGTTAAATCAGGTAAGAAAGATAAATA 11674
 QY 969 CGACAAATGGGGGTTTCCCGGTGAGCGCGGAGCTTAAGGAATCGAACCTTATCTGAA 1028
 Db 11675 TAATCATGGGATCTTAATACGAGTCAAGAAACCGGAATACCAACAATCATAGA 11734
 QY 1029 TTCAATTAAGTACCGCTCTGGGAGAAATCTCATCTGTTTACCGTGGGGCGGAGTTCA 1088
 Db 11735 TGCAAAAGTACGCAATTTCTGCCGGAATATGTAACCATCGAGGTCAATTTACGA 11794
 QY 1089 GAGCTGCTCATGAAGACGAGGAGTTCTTTCGACGACACAGGTGAA-----CTTTCG 1142
 Db 11795 TGCAGAGCTCGTATGACTCAGCCAGCGGTAAAGAAACACAGAAACACAGTCTGTTT 11854
 QY 1143 GCAGAAAGCTGTCGCTGTTTCTGAGGATGAGTGCATCTCAGGATGCACTTCGCT 1202
 Db 11855 AATTAAACAGAACTGTTTATAGAAATGAATATGACGACGAGTCTCTCGCCT 11914
 QY 1203 GACTCGCGGACGCGCTATGAACATCATGACCAATTCGGGGACACTTCAGTCCGGTGC 1262
 Db 11915 GACTCGGAGCTGCTCTCGAATATCATGAAATCTATGCGAGTTACTGGAATCCAAAT 11974
 QY 1263 ATATCTGCTGGATGTCGAGATGCTGAGCGCTGAAAGCGGTGTGACCAAGGATA 1322
 Db 11975 GTAGCTGTTTAACTGACCGGATAATCTCACACTCAAGGGGGGATGCAAAAGCA 12034
 QY 1323 TAAGGCACCCAGATGGGCGAGCTACATAAAGGATAGTGTGTTCGGGGCAGGAAA 1382
 Db 12035 TCGGCTCTCTCAATTCGTGAGGTGAGTCTCTGGATTTGGAACACTGACGCGGTGTC 12094
 QY 1383 AACAAATCTACTTGGTAACCCGACCTGAAGCCGGAAGAGAGCGTCAATATCAGGCTGG 1442
 Db 12095 CTCTATTATGATGGAACAGGACCTGAAACCGGAGACAGGTGTAAACCGAAGATCGG 12154
 QY 1443 GGTGTATTACGATAACCCCGCGCTCTGGAATGCAATGTCCAAATGTTCAGGTTTATGACTGAC 1502
 Db 12155 TATTATTATAGTAAATAGTGTGTTTTCGGCGAGCGGACGCTGTTTAAATCTGATT 12214
 QY 1503 CTCACAAAGATGCTCTCTTATCCATAAATGATAA 1538
 Db 12215 TAAATAAAGTTGACCAAGTTACGATATAGGTACAAA 12250
 RESULT 11
 ID AAQ71387 standard; cdna; 1535 BP.
 AC AAQ71387;
 XX
 DT 25-MAR-2003 (revised)
 DT 14-APR-1995 (first entry)
 XX
 DE Vibrio cholerae irgA promoter.
 XX
 XX Promoter; irgA; vaccine; cholera; iron-regulated promoter; ss.
 XX Vibrio cholerae.
 XX
 XX Key Location/Qualifiers
 FT promoter 905..1438
 FT /*tag= h
 FT promoter 905..1079
 FT /*tag= g
 FT promoter 905..1041
 FT /*tag= f
 FT promoter 922..1438
 FT /*tag= i
 FT promoter 922..1079
 FT /*tag= d
 FT promoter 922..1041
 FT /*tag= c
 FT promoter 1000..1438

FT promoter /*tag= j
 1000..1079
 FT /*tag= e
 1000..1041
 FT /*tag= b
 /note= "nts 991-1041 are claimed"
 1077..1535
 FT /*tag= a
 PN W09419482-A1.
 PD 01-SEP-1994.
 PF 22-FEB-1994; 94WO-US001780.
 XX 22-FEB-1993; 93US-00020501.
 PA (GCHO) GEN HOSPITAL CORP.
 PA (HARD) HARVARD COLLEGE.
 XX Calderwood SB, Butterson JR, Mekalanos JJ;
 DR WPI; 1994-294344/36.
 XX P-PSDB; AAR60561.
 PT New bacterial chromosomes, partic. V. cholerae, for vaccines - contg. DNA
 PT encoding a heterologous antigen linked to an iron-regulated promoter.
 XX Disclosure; Page 27-28; 44pp; English.
 CC V. cholera irgA promoter is a naturally occurring promoter which is iron-
 regulated. A DNA sequence encoding a heterologous antigen functionally
 linked to the irgA promoter permits significantly higher (ie at least ten
 fold and pref. 100 fold) expression of the heterologous antigen in a low
 iron environment, such as in an animal's intestine, than in a high-iron
 environment, such as under typical in vitro culture conditions. Pref.
 parts of the promoter sequence are indicated in the FT above. (Updated on
 25-MAR-2003 to correct PN field.)
 XX Sequence 1535 BP; 394 A; 363 C; 395 G; 383 T; 0 U; 0 Other;
 SQ
 Query Match 5.0%; Score 105; DB 2; Length 1535;
 Best Local Similarity 57.4%; Pred. No. 2e-22;
 Matches 214; Conservative 0; Mismatches 150; Indels 9; Gaps 1;
 QY 71 CAGAGGATGTGATGATGCTCTCGGCATCCGCTATGAGAAAGCTGACTAACGACGCG 130
 Db 1188 CGATGAACCATGGTGTGCTACTCGGGGGATACGGCAAGTGAATTCAAAATGACCCAG 1227
 QY 131 CCAGTGTCTGTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACACGATCTGGCGG 190
 Db 1228 CCAGTATCAGTGTGATTTCAAGAGAAGATCTGGAATCTCGCTATTACCGTATGTGACCG 1287
 QY 191 AGGCTCTGAGATCAGTAGAGGCTGTGATGTTGAAGTGTACGGTAAACCGGAGGC 250
 Db 1288 ATGCGCTAAAAGCGTACCGGGTGTACAGTACCGGAGGGGGGATACCTACC----- 1340
 QY 251 TGGAAATCAGCATCCGAGGAATGCCAGCTTACAGCTGATGATGATGATGATGATGATGAT 310
 Db 1341 --GATATCAGCATTCGTGTGATGGATCAAACTATCTTATCTTGTGTGATGATGATGATGAT 1398
 QY 311 GTACGGGGGAAGCAGTACGTGACTCCCAACGGTTTTTTCGCCATGAATACCGGTTCA 370
 Db 1399 GCCAAACCTCAGCCAGACCCCGTCCAAACAGCATGCGCCCGGCAITTCAGCAAGTTGGT 1458
 QY 371 TGCCCTCTCTGCGCCCATTCAGCGTATTGAGGTATTACGGGGGGCGGATGTCACACTGT 430
 Db 1459 TACCGCACTGCAAGCGATTGAAAGTATCGAGGTGATCCGTGGCCCGGATGTCACGCTGT 1518
 QY 431 ATGGCTCTGATGC 443
 Db 1519 ACGGCTCGGATGC 1531

CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64157-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 3675 BP; 914 A; 1051 C; 975 G; 735 T; 0 U; 0 Other;

Query Match 3.3%; Score 69.4; DB 5; Length 3675;
 Best Local Similarity 57.7%; Pred. No. 1.1e-10;
 Matches 124; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
 QY 418 ATGTCACACATGCTGATGCGATGCGCGGTGGTGAATATCATTCACGAAAG 477
 Db 1 ATGTCGTCGCTGACGGTTCGATGCGCTGCGGGGTGTAGTGAATATCATCACCACAAA 60
 QY 478 AATGACAGCAAAATGGCTCTTCGCTCAATGTCAGGGCTGAATCTGAGGAAGCAACAA 537
 Db 61 ATCGGTGAGAAATGGTCGGGTACCGTTACCGTCGATACCAACATTCAGGAACATCCGAT 120
 QY 538 TGGGGTAACAGCAGCCAGTTTATCTGAGCAGAGTGGTCCCTTGTGATGATTCGTGC 597
 Db 121 CGCGGTGACACCTATACGGTACGTTCTTTACAGTGGACCATTAATGATGATGCTGCTG 180
 QY 598 AGCCTGACGAGTACGCGGTAGCACACACACGCTCA 632
 Db 181 GGAATGAAGCTTACGGCAGCCTGGCAACAGTGA 215

RESULT 14
 AAD03054
 ID AAD03054 standard; DNA; 2495 BP.
 XX
 AC AAD03054;
 XX
 DT 13-JUN-2001 (first entry)
 XX
 DE ironNec gene fragment from an extraintestinal isolate of Escherichia coli.
 XX
 KW ironNec; extracytoplasmic protein; immunogen; vaccine; BI; UTI;
 XX immunotherapy; extraintestinal infection; urinary tract infection;
 KW meningitis; pneumonia; intra-abdominal infection; antibiotic; ds.
 XX
 OS Escherichia coli.
 XX
 XX Key Location/Qualifiers
 XX CDS 321..2495
 XX /*tag= a
 XX /product= "E. coli ironNec extracytoplasmic protein"
 XX /transl_except= (pos:678..686, aa:Val-Tyr)
 XX /note= "Insertion of 3 bases alters the reading frame"
 XX /transl_except= (pos:723..728, aa:Asn-Arg-Tip)
 XX /note= "There is an apparent deletion of one codon
 XX corresponding to Arg, which alters the reading frame; the
 XX coding region does not include stop codon"
 XX /partial
 XX 321..392
 XX /*tag= b
 XX 393..2495
 XX /*tag= c
 XX /product= "Mature E. coli ironNec extracytoplasmic
 XX
 FT sig_peptide
 FT mat_peptide

FT protein"
 XX WO2000121636-A1.
 PN 29-MAR-2001.
 PD 22-SEP-2000; 2000WO-US026117.
 XX 22-SEP-1999; 99US-0155621P.
 PR (UUNY) UNIV NEW YORK STATE RES FOUND.
 XX Russo T, Carlino U;
 PI MPI; 2001-244936/25.
 DR P-PSDB; AAY72914.
 XX Novel isolated ironNec polynucleotide from extraintestinal isolate of
 PT Escherichia coli useful as vaccine for treating or preventing
 PT extraintestinal infections caused by extraintestinal pathogenic
 PT Escherichia coli.
 XX Claim 5; Page 37-38; 44pp; English.
 CC The invention relates to the identification of ironNec gene, from an
 CC extraintestinal isolate of Escherichia coli. This gene is expressed in
 CC increased amounts in human urine and is identified by transposon (TnphoA)
 CC mutagenesis. ironNec gene encodes an extracytoplasmic protein. This gene
 CC can be used as an immunogen in vaccine formulations. The recombinant
 CC vector comprising nucleotide sequence encoding one or more antigenic
 CC epitope of ironNec is useful for diagnostic and immunotherapeutic
 CC purposes. The ironNec antigenic peptide is useful for treating or
 CC preventing extraintestinal infections (BIs) caused by extra- intestinal
 CC pathogenic E. coli (ExPEC). The BIs include urinary tract infection
 CC (UTI), meningitis, intra-abdominal infection and pneumonia. The present
 CC sequence is ironNec gene, from an extraintestinal isolate of Escherichia
 CC coli
 XX
 SQ Sequence 2495 BP; 686 A; 557 C; 632 G; 620 T; 0 U; 0 Other;
 Query Match 3.3%; Score 68.6; DB 4; Length 2495;
 Best Local Similarity 55.1%; Pred. No. 1.1e-10;
 Matches 134; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
 QY 376 CCTCTGCGCCCATGTAGCGGTATTCAGGGGGCGGATGTCACACTGTATGGC 435
 Db 732 CCACCGCAACAGTTGAGCGGTATTCAGCGGCGCTCGCGGGCGCGGTACGGT 791
 QY 436 TCTGATGCGATGGCGGTGTGTCATATCATTTACCAAGAAAGATGCAGACAAATGGCTC 495
 Db 792 TCGGGGGCGCGGGGGGTGTGTAACATCATTTACCAACGTCGCCACACGACTGGCAC 851
 QY 496 TCTTCCCGTCAATGCAGGGCTGAATCTGCAGAAAGCAACAAATGGGTACAGCAGCGAG 555
 Db 852 GGTTCGCTGTGCTTATACACCAACAGCCGCGAAAGTAGCGAAGAGGGCGGTACCGCTGC 911
 QY 556 TTTAATTTCTGAGCAGTGGTCCCTTGTGATCATCTGTACGCTGAGGTACGGGT 615
 Db 912 GCCAATTTTCAGCCTTAGTGGGCGCTCTGGCTGGTGTCTTTACACGCGGTTGTATGGT 971
 QY 616 AGC 618
 Db 972 AAC 974
 XX
 XX RESULT 15
 XX AAL15184/c
 XX ID AAL15184 standard; DNA; 4039 BP.
 XX AAL15184;
 XX AC AAL15184;
 XX 04-SEP-2000 (first entry)
 DT
 XX

DE DNA encoding Escherichia coli virulence proteins.

KW Virulence protein; tatA; tatB; tatC; tatE; mdoG; creC; recG; yggN; eck1;
KW iroD; iroC; iroE; mtd2; msl1; vaccine; infection; Gram negative Bacterium;
KW ss.

XX Escherichia coli.

XX Key Location/Qualifiers

XX CDS 1..285

XX /tag= a

XX /product= "virulence protein"

XX /note= "encodes AAY93235"

XX CDS 370..1326

XX /tag= b

XX /product= "virulence protein"

XX /note= "encodes AAY93236"

XX WO200028038-A2.

XX PD 18-MAY-2000.

XX XX

XX PF 09-NOV-1999; 99WO-GB003721.

XX XX

XX PR 09-NOV-1998; 98GB-00024569.

XX PR 09-NOV-1998; 98GB-00024570.

XX PR 17-DEC-1998; 98GB-00027814.

XX PR 17-DEC-1998; 98GB-00027815.

XX PR 17-DEC-1998; 98GB-00027816.

XX PR 17-DEC-1998; 98GB-00027818.

XX PR 13-JAN-1999; 99GB-00000708.

XX PR 13-JAN-1999; 99GB-00000710.

XX PR 13-JAN-1999; 99GB-00000711.

XX PR 28-JAN-1999; 99GB-00001915.

XX XX (MICR-) MICROSCIENCE LTD.

XX PA

XX PI Crooke HR, Clarke EE, Everest PH, Dougan G, Holden DW, Shea JE;

XX PI Feldman RG;

XX XX

XX WPI; 2000-376550/32.

XX DR P-PSDB; AAY93235, AAY93236.

XX PT Peptide encoded by an operon including genes from Escherichia coli for
XX screening potential drugs, detecting virulence and treating conditions
XX associated with infection by a Gram negative bacterium.

XX PS Example 10; Page 74-78; 122pp; English.

XX CC The present sequence encodes Escherichia coli virulence proteins. The
XX specification describes virulence proteins which are encoded by an operon
XX including tatA, tatB, tatC, tatE, mdoG, recC, recG, yggN, eck1, iroD,
XX iroC, iroE, mtd2 or msl-16 genes obtained from Escherichia coli K1. The
XX virulence proteins and polynucleotides, and their vaccines are useful for
XX screening potential drugs, for the detection of virulence, and for
XX treating or preventing conditions associated with infection by a Gram
XX negative bacterium particularly Escherichia coli

XX SQ Sequence 4039 BP; 985 A; 1039 C; 942 G; 1073 T; 0 U; 0 Other;

XX Query Match 3.2%; Score 66.6; DB 3; Length 4039;

XX Best Local Similarity 54.8%; Pred. No. 9.4e-10;

XX Matches 132; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 376 CCTCTGGCGCCATTGAGCGTATTGAGTTATCAGGGGCGCGATGCCACACTGTATGGC 435

Db 3138 CCACCGGAACAGGTTGAGCGTATTGAGTTATCAGGGGCGCGATGCCACACTGTATGGC 3079

QY 436 TCTGATCGGATGGCGGCGTGTGGTGAATATCATTACCAAGAAATCCAGACAAATGGCTC 495

Db 3078 TCGGGGGACCGCGGGGGGTGGTGAACATCATTACCAACGTCGCCACCAACCACTGGCAC 3019

QY 496 TCTTCGTCATGACAGGCGTGAATCTGCAGGAAAGCAAAATGGGGTAAACAGAGCCAG 555

Db 3018 GGTTCGCTGTGTTATACCAATCAGCCGAAAGTAGCGATGAGGGCGCTACCGCTCGC 2959

QY 556 TTTAATTTCTGGACAGTGGTCCCTTTGTGATGATTCTGTACGCTGCAGGTACGGGT 615

Db 2958 GCCAATTTACGCTTAGTGGGCTCTGGCTGGTAAATGCTCTTACCAACGCTTTGTATGGT 2899

QY 616 A 616

Db 2898 A 2898

Search completed: October 12, 2004, 12:10:32
Job time : 826 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 12, 2004, 11:56:38 ; Search time 8116 Seconds
(without alignments)
11166.866 Million cell updates/sec

Title: US-10-625-972-4
Perfect score: 2091
Sequence: 1 atcggaataacactctggc.....cgctgaactatcagttctga 2091

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.hg.hum.*
31: em.hg.inv.*
32: em.hg.other.*
33: em.hg.mus.*
34: em.hg.pln.*
35: em.hg.rod.*
36: em.hg.nam.*
37: em.hg.vit.*
38: em.sy.*
39: em.hgo.hum.*
40: em.hgo.mus.*
41: em.hgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2091	100.0	2091	6	AR025458	Sequence
2	2091	100.0	8041	6	AR025455	Sequence
3	2089.4	99.9	8040	1	AF126104	Escherich
C 4	2089.4	99.9	10029	1	AE005310	Escherich
C 5	2089.4	99.9	10155	1	AE005274	Escherich
C 6	2089.4	99.8	87583	6	AE204161	Sequence
C 7	2087.8	99.8	86248	6	BD184766	Nucleic a
C 8	2087.8	99.8	327773	1	AP002554	Escherich
C 9	2079.8	99.5	76804	6	AX702523	Sequence
C 10	2079.8	99.5	303121	1	AF081285	Escherich
C 11	2046.8	97.9	13710	1	AF081285	Escherich
C 12	1863.8	89.1	37710	1	ECO278144	Escherich
C 13	1025	49.0	71684	1	AF447814	Escherich
C 14	646.4	30.9	10029	1	AE013687	Yersinia
C 15	644.8	30.8	220050	1	AJ414156	Yersinia
C 16	177	8.5	343473	1	EX640451	Bordetella
C 17	175.4	8.4	3169	1	BBU56084	Yersinia
C 18	174.4	8.3	300507	1	AE017146	Helicobac
C 19	173.8	8.3	1239	1	AY212670	Unculture
C 20	153.8	7.4	317511	1	CJ11168X3	Campyloba
C 21	150	7.2	3300	1	CCU80812	Campylobact
C 22	149.4	7.1	345829	1	EX571868	Photornab
C 23	149.4	7.1	349980	6	AX770905	Sequence
C 24	147.8	7.1	10354	1	AE013889	Yersinia
C 25	147.8	7.1	199050	1	AJ414147	Yersinia
C 26	142.6	6.8	312839	1	AE016858	Pseudomon
C 27	137.8	6.6	12635	1	AE009322	Agrobacte
C 28	137.8	6.6	13095	1	AE008292	Agrobacte
C 29	134.2	6.4	3339	1	AE052655	Vibrio pa
C 30	134.2	6.4	296650	1	AP005082	E.coli coli
C 31	130.6	6.2	2499	1	ECOCIR	Escherich
C 32	130.6	6.2	10342	1	AE005447	Escherich
C 33	130.6	6.2	10353	1	AE000304	Escherich
C 34	130.6	6.2	10353	6	AX370264	Sequence
C 35	130.6	6.2	75888	1	ECOHU47	Sequence
C 36	130.6	6.2	268857	1	AP002560	Escherich
C 37	129	6.2	12697	1	AE015239	Shigella
C 38	129	6.2	295689	1	AE016985	Shigella
C 39	127	6.1	2055	6	AR375730	Sequence
C 40	125.6	6.0	2046	6	AX276467	Sequence
C 41	125.6	6.0	2046	6	AX844854	Sequence
C 42	125.6	6.0	2046	6	AX828804	Sequence
C 43	125.6	6.0	2049	1	AF320691	Escherich
C 44	125.6	6.0	12943	6	AX276440	Sequence
C 45	125.6	6.0	306118	1	AE016771	Escherich

ALIGNMENTS

RESULT 1
AR025458
LOCUS AR025458 2091 bp DNA linear PAT 05-DEC-1998
DEFINITION Sequence 4 from patent US 5798260.
ACCESSION AR025458
VERSION AR025458.1 GI:3978086
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 2091)
AUTHORS Tarr, P.I., Bilge, S.S., Besser, T.E. and Vary, J.C. Jr.
TITLE Escherichia coli O157:H7 epithelial adhesin
JOURNAL Patent: US 5798260-A 4 25-AUG-1998;
FEATURES Location/Qualifiers

Pred. No. is the number of results predicted by chance to have a


```

source
1. .2091
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 2091; DB 6; Length 2091;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2091; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATCGGAATAACCACTCTCGCTCCGTAGTCAATCCCTGTCTCGGATTTTCAGCAGCAGC 60
Db 1 ATCGGAATAACCACTCTCGCTCCGTAGTCAATCCCTGTCTCGGATTTTCAGCAGCAGC 60

Qy 61 ATAGCTGTCGACAGAGATGATGATTGTCTCGGCATCCGGCTATCAGAAAAAGCTGACT 120
Db 61 ATAGCTGTCGACAGAGATGATGATTGTCTCGGCATCCGGCTATGAGAAAAGCTGACT 120

Qy 121 AACGACGCCCGAGTGTTCCTGTGATTAGCCAGAGGAATTGCAGTCCAGCCAGTACCAC 180
Db 121 AACGACGCCCGAGTGTTCCTGTGATTAGCCAGAGGAATTGCAGTCCAGCCAGTACCAC 180

Qy 181 GATCTGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGATGTTCAAAAGTGGTAGGGTAAA 240
Db 181 GATCTGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGATGTTCAAAAGTGGTAGGGTAAA 240

Qy 241 ACCGAGGGCTCGAAATCAGCATCCGAGGAATGCCAGCCAGTTTACCGCTGATCTGATT 300
Db 241 ACCGAGGGCTCGAAATCAGCATCCGAGGAATGCCAGCCAGTTTACCGCTGATCTGATT 300

Qy 301 GATGGTTCGTGAGGGCGGAGCAGTGACGTGACTCCCAACGGTTTTTCTGCCATGAAT 360
Db 301 GATGGTTCGTGAGGGCGGAGCAGTGACGTGACTCCCAACGGTTTTTCTGCCATGAAT 360

Qy 361 ACCGGGTTCAATCCGCCCTCTGCGGCCCATTTGAGCGTATTGAGGTTATCAGGGGGCCGATG 420
Db 361 ACCGGGTTCAATCCGCCCTCTGCGGCCCATTTGAGCGTATTGAGGTTATCAGGGGGCCGATG 420

Qy 421 TCACACTGTATGGCTCTGATGCGATGGGCGGTGTGGTGAATATCATTACACAGAAGAT 480
Db 421 TCACACTGTATGGCTCTGATGCGATGGGCGGTGTGGTGAATATCATTACACAGAAGAT 480

Qy 481 GCAGACAAATGGCTCTCTTCCGTCAATGCAGGGCTGAATCTGCAGGAAAGCAACAAATGG 540
Db 481 GCAGACAAATGGCTCTCTTCCGTCAATGCAGGGCTGAATCTGCAGGAAAGCAACAAATGG 540

Qy 541 GGTAAACAGCAGCCAGTTTAAATTTCTGAGCAGTGGTCCCTTGTGATGATTTCTGTACG 600
Db 541 GGTAAACAGCAGCCAGTTTAAATTTCTGAGCAGTGGTCCCTTGTGATGATTTCTGTACG 600

Qy 601 CTGCAGGTACCGGCTAGCACACAACAGCGTTCAGGGTTTCATCGGTACATCATCTGAGCGAT 660
Db 601 CTGCAGGTACCGGCTAGCACACAACAGCGTTCAGGGTTTCATCGGTACATCATCTGAGCGAT 660

Qy 661 ACAGCAGGACCGGTATTTCTTTATCCACGAGATCAGAAATTTAATCTTGGTGCAGT 720
Db 661 ACAGCAGGACCGGTATTTCTTTATCCACGAGATCAGAAATTTAATCTTGGTGCAGT 720

Qy 721 CTTGACTGGAAGCGTCCGAGCAGGATGTGCTCTGGTTTGATATGATACCAACCCGGCAG 780
Db 721 CTTGACTGGAAGCGTCCGAGCAGGATGTGCTCTGGTTTGATATGATACCAACCCGGCAG 780

Qy 781 CGTTATGATAACCGGGATGGGCAACTGCGGAGTCTGACGGGGGATATGACCGGACCCCTG 840
Db 781 CGTTATGATAACCGGGATGGGCAACTGCGGAGTCTGACGGGGGATATGACCGGACCCCTG 840

Qy 841 CGCTATGAGCGAAAACAAATTTTCACTTGGCTATGATCATACTTTTCACTTCGGAACAATGG 900
Db 841 CGCTATGAGCGAAAACAAATTTTCACTTGGCTATGATCATACTTTTCACTTCGGAACAATGG 900

Qy 901 AAATCGTATCTGAACTGGAAACGACAGCAAAATTAAGTCTGTGAGCTTTGATCGCAGTGTA 960
Db 901 AAATCGTATCTGAACTGGAAACGACAGCAAAATTAAGTCTGTGAGCTTTGATCGCAGTGTA 960

```



```

Db      4776 ACCGAGAGGTGGCATCGCTGGTGGTCCGTTATCGCGGGAACACCAACGTTTCACC 4835
Qy      1801 CAGAAATATTGTCACATGACGCTGTACAGAAAGAGTATGATGAGAAAGGAGAAATAC 1860
Db      4836 CAGAAATATTGTCACATGACGCTGTACAGAAAGAGTATGATGAGAAAGGAGAAATAC 4895
Qy      1861 CTGAAAGCCTGGAGCGGTGGATGAGGTCCTGCTGGGAAGATGACGATGCCCTGACG 1920
Db      4896 CTGAAAGCCTGGAGCGGTGGATGAGGTCCTGCTGGGAAGATGACGATGCCCTGACG 4955
Qy      1921 CTGAATCTCGGTGTAATACCTGCTCAACAAGGATTACAGTGACGTGACGCTGTACAGT 1980
Db      4956 CTGAATCTCGGTGTAATACCTGCTCAACAAGGATTACAGTGACGTGACGCTGTACAGT 5015
Qy      1981 GCCGTTAAGTAGTACGCTGTATGCGGTGATTCTTCAGACGGGATCATCAACACAGGA 2040
Db      5016 GCCGTTAAGTAGTACGCTGTATGCGGTGATTCTTCAGACGGGATCATCAACACAGGA 5075
Qy      2041 TATGTGATACCTGACGCAAAATTACTGTGATGCTGCTGAATCATCAGTTCTGA 2091
Db      5076 TATGTGATACCTGACGCAAAATTACTGTGATGCTGCTGAATCATCAGTTCTGA 5126

RESULT 3
AF126104
LOCUS      8040 bp      DNA      linear      BCT 12-MAR-2002
DEFINITION Escherichia coli O157:H7 adhesin (iha), putative tellurium
            resistance protein A (tlrA), putative tellurium resistance protein
            B (tlrB), putative tellurium resistance protein C (tlrC), and
            putative tellurium resistance protein D (tlpD) genes, complete cds.
ACCESSION AF126104
VERSION    AF126104.2
KEYWORDS   GI:19352330
SOURCE     Escherichia coli O157:H7
ORGANISM   Escherichia coli O157:H7
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE  1 (bases 1 to 8040)
            Tarr,P.I., Bilge,S.S., Vary,J.C. Jr., Jelacic,S., Habeeb,R.L.,
            Ward,T.R., Baylor,M.R. and Besser,T.E.
            Iha: a novel Escherichia coli O157:H7 adherence-conferring molecule
            encoded on a recently acquired chromosomal island of conserved
            structure
JOURNAL    Infect. Immun. 68 (3), 1400-1407 (2000)
MEDLINE    20143752
PUBMED     10678953
REFERENCE  2 (bases 1 to 8040)
            Tarr,P.I., Bilge,S.S., Vary,J.C., Jelacic,S., Habeeb,R.L.,
            Ward,T.R. and Baylor,M.
            Direct Submission
            Submitted (04-FEB-1999) Pediatrics/Gastroenterology, Children's
            Hospital and Regional Medical Center, 4800 Sand Point Way NE,
            Seattle, WA 98105, USA
            3 (bases 1 to 8040)
            Tarr,P.I., Bilge,S.S., Vary,J.C., Jelacic,S., Habeeb,R.L.,
            Ward,T.R. and Baylor,M.
            Direct Submission
            Submitted (12-MAR-2002) Pediatrics/Gastroenterology, Children's
            Hospital and Regional Medical Center, 4800 Sand Point Way NE,
            Seattle, WA 98105, USA
            Sequence update by submitter
REMARK     On Mar 12, 2002 this sequence version replaced gi:7108479.
FEATURES   Location/Qualifiers
            source
            1..8040
                /organism="Escherichia coli O157:H7"
                /mol_type="genomic DNA"
                /strain="86-24"
                /serotype="O157:H7"
                /db_xref="taxon:83334"
            /note="KpnI fragment; portion of tellurite resistance and
            adherence-conferring island (TAI)"
            3036..5126
                /gene="iha"

```

```

3036..5126
/gene="iha"
/function="confers upon nonadherent laboratory Escherichia
coli the ability to adhere to epithelial cells in vitro"
/note="similar to IrgA"
/codon_start=1
/transl_table=11
/product="adhesin"
/protein_id="AAF36432.2"
/db_xref="GI:19352331"
/translation="MRITLASVVPICLGSFASIAAEVDMLVSAGYKKLTNAAA
SVYSIQEELSSQYHDLAALRSVEGVDSCTGKTGLEISIRGMPASYTILIDG
VROGSSDVTNPNGPFSAMTGTMPPLAIEIRIEVIRGPMSTLYGSDAMGVNITRN
ADKLSSVAGNLQESNKWGNSSQNFWSGGLVDDSVSLQVRGSGTQORQSGSVTSL
SDTAATPIPTSSQYNLCAEALDKASQDVLWEDMDTTRQYDNDRDGLSLTGY
DRLIRYRNKISAGYDHTFTFGTWKSYLNWNETENKRELVRSVLKRDKWGLAGQPRE
LKENLILNLSLLTLPGLSEHLVTVGGEFQSSNMKGDVLLASTGETFRQKSWSPAEDE
WHLTDALTAGSRYEHEQFGHFSPRAYLVMDVADATLKGGVTTGYKAPRMGQLH
KGISGVSQGTNLLGNPDLKPEESVSEAGVYDNPAGLNANVTQDFMDFSNKIVSY
KISDNTNSYVNSGKARLHGVFACTLPWSEDTLSLNTYTWTRSEQDGDNGKAPLSY
TPHMVAKLNWQITEVASWLGARVGTPTFTQNYSSLSAVQKQVYDEKGYLKAW
TVVDAGLSWKMTDALTLNAAVNNLLNKDYSVLSAGKSTLYAGDYFQTGSGSTGVV
IPERNYMSLNYQP"
complement(5718..6026)
/gene="tlrA"
complement(5718..6026)
/gene="tlrA"
/note="similar to antimicrobial-resistance protein;
similar to TsrD of Alcaligenes sp.: SwissProt Accession
Number P18781"
/codon_start=1
/transl_table=11
/product="putative tellurium resistance protein A"
/protein_id="AAF36433.1"
/db_xref="GI:7108481"
/translation="MNLQSGNPLQOQTIRNLQYPAKSGFKGEPDTCFLMNAOGK
VSGSDSDFIFNNLSSPEGAVRLVTGQQAISIEALDRVPANVSKIALTVVVDGDTIS
"
complement(6448..7023)
/gene="tlrB"
complement(6448..7023)
/gene="tlrB"
/note="similar to antimicrobial-resistance protein;
similar to TsrE of Alcaligenes sp.: SwissProt Accession
Number P18782"
/codon_start=1
/transl_table=11
/product="putative tellurium resistance protein B"
/protein_id="AAF36434.1"
/db_xref="GI:7108482"
/translation="MAVSLYKGGNLSLTKKAPTMYAVMVLGNDABVTGCGFDLDAS
KVLVFAVTIYDAEARKQNFQWNSFMRVYNNNDNGTEIARFDLSEDASTEAMVFGEL
YRHGTWKFKAVGQGFAGLSALASQHGYNV"
complement(7092..7670)
/gene="tlrC"
complement(7092..7670)
/gene="tlrC"
/note="similar to antimicrobial-resistance protein;
similar to TsrD of Alcaligenes sp.: SwissProt Accession
Number P18781"
/codon_start=1
/transl_table=11
/product="putative tellurium resistance protein C"
/protein_id="AAF36435.1"
/db_xref="GI:7108483"
/translation="MVSVSTGGNLSLSKAAPMKNVVLGLDWABSTGCGFDLDAS
AFLIANGKVRGDSDFIFNNLTSDDGSVTHTDNRGTGEGDDEQVKLIDTKVASDI
KDLIFVVTIHDQAARQSFQVSGAFIRLVNDNDNQTEVARIYDLTDEASTAMLPGEL
YRHNGEMKFAVGGVAGGLASVCAQYGINAS"
complement(7719..7988)
/gene="tlpD"
complement(7719..7988)

```



```

/genes="t1pD"
/note="similar to antimicrobial-resistance protein;
similar to TerC of Alcaligenes sp.: SwissProt Accession
Number P18780"
Number P18780
/codon_start=1
/transal_table=11
/product="putative tellurium resistance protein D"
/protein_id="AAP36436.1"
/db_xref="GI:7108484"
/translation="NMFAIGLRTLVFVLEAKGVKLEKAVILVFFFAAKLGINA
TDIWHGVSIAATSTLYVLVGLAIGLGSAMVPEGESEKGR"

```

ORIGIN

Query Match.	99.9%;	Score	2089.4;	DB	1;	Length	8040;
Best Local Similarity	100.0%;	Prod.	No. 0;				
Matches	2090;	Conservative	0;	Mismatches	1;	Indels	0;
							Gaps
							0;
Qy	1	ATCGGAATAACCACTCTGGCTTCGGTAGTCATTTCCCTGTCTCGGATTTTTCAGCCAGCAGC	60				
Db	3036	ATCGGAATAACCACTCTGGCTTCGGTAGTCATTTCCCTGTCTCGGATTTTTCAGCCAGCAGC	3095				
Qy	61	ATAGCTCTGCAGAGGATGTGATGTGTCGGCATTCGGCTATGAGAAAAAGCTTGACT	120				
Db	3096	ATAGCTCTGCAGAGGATGTGATGTGTCGGCATTCGGCTATGAGAAAAAGCTTGACT	3155				
Qy	121	AACGACGCCAGCTGTTCTGTGATTAGCCAGAGGAATTCAGTCACGCCAGTACCAC	180				
Db	3156	AACGACGCCAGCTGTTCTGTGATTAGCCAGAGGAATTCAGTCACGCCAGTACCAC	3215				
Qy	181	GATCTGCGGAGGCTCTGAGATCAGTAGAGGGGTGTGGATTTGAAAGTGTAACGGGTAAA	240				
Db	3216	GATCTGCGGAGGCTCTGAGATCAGTAGAGGGGTGTGGATTTGAAAGTGTAACGGGTAAA	3275				
Qy	241	ACCGAGGGCTGGAATCAGATCCGAGGAATGCCAGCCAGTTACACGCTGATCTGATT	300				
Db	3276	ACCGAGGGCTGGAATCAGATCCGAGGAATGCCAGCCAGTTACACGCTGATCTGATT	3335				
Qy	301	GATGGTGTTCGTTCAGGCGGGAAGCAGTGACGTGACTCCCAACGGTTTTTCTGCCATGAAT	360				
Db	3336	GATGGTGTTCGTTCAGGCGGGAAGCAGTGACGTGACTCCCAACGGTTTTTCTGCCATGAAT	3395				
Qy	361	ACCGGGTTTCATGCCCTCTTGATGCGATGGCGCCGATTCAGGCTATCAGGGGGCCGATG	420				
Db	3396	ACCGGGTTTCATGCCCTCTTGATGCGATGGCGCCGATTCAGGCTATCAGGGGGCCGATG	3455				
Qy	421	TCCACACTGTATGCTCTGTATGCGATGGGCGGTCTGGTGAATATCATACAGAAAGAAT	480				
Db	3456	TCCACACTGTATGCTCTGTATGCGATGGGCGGTCTGGTGAATATCATACAGAAAGAAT	3515				
Qy	481	GCAGACAAATGGCTCTCTTCGGTCAATTCAGGGCTGAATTCGACGGAAGCAACAAATGG	540				
Db	3516	GCAGACAAATGGCTCTCTTCGGTCAATTCAGGGCTGAATTCGACGGAAGCAACAAATGG	3575				
Qy	541	GGTAACAGCAGCCAGTTTAATTTCTGGACAGTTCGTCCTCTTGGATGATTTCTGTCCAG	600				
Db	3576	GGTAACAGCAGCCAGTTTAATTTCTGGACAGTTCGTCCTCTTGGATGATTTCTGTCCAG	3635				
Qy	601	CTGCAGGTACCGGTAGCACACACACGCTCAGGGTTTCATCGGTACATCAGTCAGCGAT	660				
Db	3636	CTGCAGGTACCGGTAGCACACACACGCTCAGGGTTTCATCGGTACATCAGTCAGCGAT	3695				
Qy	661	ACAGCAGCAGCGGTATTCCTTTATCCACGAGTTCACAGATTAATATCTTGGTGACGT	720				
Db	3696	ACAGCAGCAGCGGTATTCCTTTATCCACGAGTTCACAGATTAATATCTTGGTGACGT	3755				
Qy	721	CTTGACTCGAAGGCGTCGGAGCAGGATGTGCTCTGTGTTGATATGGATACACCCCGCAG	780				
Db	3756	CTTGACTCGAAGGCGTCGGAGCAGGATGTGCTCTGTGTTGATATGGATACACCCCGCAG	3815				
Qy	781	CGTTATGATAACCGGGATGGGCAACTGGGGAGTCTGACGGGGGGATATGACCGGACCCCTG	840				
Db	3816	CGTTATGATAACCGGGATGGGCAACTGGGGAGTCTGACGGGGGGATATGACCGGACCCCTG	3875				

Qy	841	CGCTATGAGCGAAACAAAATTTTCAGCTGCGCTATGATCATCTTTACCTTCGGAAACATCG	900
Db	3876	CGCTATGAGCGAAACAAAATTTTCAGCTGCGCTATGATCATCTTTACCTTCGGAAACATCG	3935
Qy	901	AAATCGTATCTGAATCGGAACGAGACAGAAAAATAAGGTCGTGAGCTTGTACGCACTGTA	960
Db	3936	AAATCGTATCTGAATCTGGAACGAGACAGAAAAATAAGGTCGTGAGCTTGTACGCACTGTA	3995
Qy	961	CTGAAGCGGACAAAATGGGGGCTTGCCGTCAGCCCGGGAGCTTAAAGAAATCGAACCTT	1020
Db	3996	CTGAAGCGGACAAAATGGGGGCTTGCCGTCAGCCCGGGAGCTTAAAGAAATCGAACCTT	4055
Qy	1021	ATCCTGAAATTCATTACTGCTTACCCCTCTGGGAGAAATCTCATCTGCTTACGTTGGGGGC	1080
Db	4056	ATCCTGAAATTCATTACTGCTTACCCCTCTGGGAGAAATCTCATCTGCTTACGTTGGGGGC	4115
Qy	1081	GAGTTTCAGAGCTCGTCCATGAAAGACGAGTTGTCCTTGCCAGCACAGGTGAAACTTTTC	1140
Db	4116	GAGTTTCAGAGCTCGTCCATGAAAGACGAGTTGTCCTTGCCAGCACAGGTGAAACTTTTC	4175
Qy	1141	CGGAGAAAGCTGGTTCGTTATTTGCTGAGGATGAGTGGCATCTCCGGATGCACATTGCG	1200
Db	4176	CGGAGAAAGCTGGTTCGTTATTTGCTGAGGATGAGTGGCATCTCCGGATGCACATTGCG	4235
Qy	1201	CTGACTCGGGCAGCCGCTATGAACATCATGACAAATTCGGGGACACTTCACTCCGCGT	1260
Db	4236	CTGACTCGGGCAGCCGCTATGAACATCATGACAAATTCGGGGACACTTCACTCCGCGT	4295
Qy	1261	GCATATCTGCTGTGGATGTGGCAGATGCTGCAGCGTGAAAGCGGTGTGCCACGGGA	1320
Db	4296	GCATATCTGCTGTGGATGTGGCAGATGCTGCAGCGTGAAAGCGGTGTGCCACGGGA	4355
Qy	1321	TATAAGGCACCCAGAAATGGGCGAGCTACATAAAGGGATAGTGGTGTGCCGGCAGGGA	1380
Db	4356	TATAAGGCACCCAGAAATGGGCGAGCTACATAAAGGGATAGTGGTGTGCCGGCAGGGA	4415
Qy	1381	AAAAAATCTACTTGGTAAACCCCGACCTGAAACCGGAGAGAGCGTCAGTTATGAGCT	1440
Db	4416	AAAAAATCTACTTGGTAAACCCCGACCTGAAACCGGAGAGAGCGTCAGTTATGAGCT	4475
Qy	1441	GGGTGTATTACGATTAACCCCGCGGTCTCAATGCCAATGTCAAGGTTTTATGACTGAC	1500
Db	4476	GGGTGTATTACGATTAACCCCGCGGTCTCAATGCCAATGTCAAGGTTTTATGACTGAC	4535
Qy	1501	TTCTCCAAAGATTTGTCTCTTATTCATAAATGATAACCAATAGCTATGTAAACAGC	1560
Db	4536	TTCTCCAAAGATTTGTCTCTTATTCATAAATGATAACCAATAGCTATGTAAACAGC	4595
Qy	1561	GGAAAGCCCGGTGACAGGTGTGGAATTTGCCGGCACATTTGCCGTGTGCTCAGAGAT	1620
Db	4596	GGAAAGCCCGGTGACAGGTGTGGAATTTGCCGGCACATTTGCCGTGTGCTCAGAGAT	4655
Qy	1621	GTACAGCTGTCACTGAAATTAACCTGACCCGGAAGTGAAACAACGTGATGTGATAACAA	1680
Db	4656	GTACAGCTGTCACTGAAATTAACCTGACCCGGAAGTGAAACAACGTGATGTGATAACAA	4715
Qy	1681	GGTGGCGCGTGTGATTAATCCCTGAACACATGGTGAATCGGAACTGAACTGCAGATC	1740
Db	4716	GGTGGCGCGTGTGATTAATCCCTGAACACATGGTGAATCGGAACTGAACTGCAGATC	4775
Qy	1741	ACCGAAGAGGTGGCATCATGGCTGGGTGCCGTTATCCGGGAAAAACACACGTTTCACC	1800
Db	4776	ACCGAAGAGGTGGCATCATGGCTGGGTGCCGTTATCCGGGAAAAACACACGTTTCACC	4835
Qy	1801	CAGAATATTTCGTCACTGAGCGCTGTACAGAAAGATGTATGATGAGAAGGGAATAC	1860
Db	4836	CAGAATATTTCGTCACTGAGCGCTGTACAGAAAGATGTATGATGAGAAGGGAATAC	4895
Qy	1861	CTGAAGCCTGGAACGGTGGTGGATGACAGGTCGTCTGTGGAAGATGACCGATGCCCTGACG	1920
Db	4896	CTGAAGCCTGGAACGGTGGTGGATGACAGGTCGTCTGTGGAAGATGACCGATGCCCTGACG	4955
Qy	1921	CTGAATGCTGCGGTGAATAACCTGCTCAACAGGATTAAGTGAAGTGAAGTGTACAGT	1980


```

/function="orf; Adaptations, atypical conditions"
/notes="Residues 1 to 192 of 192 are 100.00 pct identical
to residues 1 to 192 of 192 from GenPept 118 :
gi|950682|gb|AA86850.1| [U59239] putative TerD [Serratia
marcescens]"
/codon_start=1
/transl_table=11
/product="putative phase inhibition, colicin resistance
and tellurite resistance protein"
/protein_id="AAG55729.1"
/db_xref="GI:12514500"
/translations="MSVLSKGGNVLSLSKAAPSKMVLVGLGWDARSTDGQDFDLSDAS
AFILASGKRVGSDDFIFNNLTSSDGSVTHGTGNDRTGGDGDSDSLKIKLDAVPSEV
DKLIFVYTHDAARQSCFQVSGSARILRVNDNQIEVARYDLTSDASTETAMLFGEI
YRHNWKFRVAGVGAGGLASVCAQYGINAS"
4167..4742
/gene="terE_2"
/notes="terE_2"
/notes="synonym: Z1615"
4167..4742
/gene="terE_2"
/functions="Orf; Adaptations, atypical conditions"
/notes="Residues 1 to 191 of 191 are 100.00 pct identical
to residues 1 to 191 of 191 from GenPept 118 :
gi|710848|gb|AAF3643.1|AF126104_3 (AF126104) putative
tellurium resistance protein B [Escherichia coli O157:H7]"
/codon_start=1
/transl_table=11
/product="putative phase inhibition, colicin resistance
and tellurite resistance protein"
/protein_id="AAG55730.1"
/db_xref="GI:12514501"
/translations="MAVSLVKGNGVSLTKEAFTMVNVVGLGWDARVTDGQDFDLSDAS
VFVAGBGRVLSDAHIFFNKTSPPGAVEHQDNRTGGDGDSDSLKIKLDAVPSEV
KELVAVIYDAARQSCFQVSGSARILRVNDNQIEVARYDLTSDASTETAMLFGEI
YRHNWKFRVAGVGAGGLASVCAQYGINAS"
5164..5472
/gene="terF_2"
/notes="terF_2"
/notes="synonym: Z1616"
5164..5472
/gene="terF_2"
/functions="Orf; Adaptations, atypical conditions"
/notes="Residues 1 to 102 of 102 are 100.00 pct identical
to residues 1 to 102 of 102 from GenPept 118 :
gi|710848|gb|AAF3643.1|AF126104_2 (AF126104) putative
tellurium resistance protein A [Escherichia coli O157:H7]"
/codon_start=1
/transl_table=11
/product="partial putative phase inhibition, colicin
resistance and tellurite resistance protein"
/protein_id="AAG55731.1"
/db_xref="GI:12514502"
/translations="MNLQSGQNPLOQSTIRNLQVPAKSGFKGEPDTCLEFMLNAQKG
VSGDSDFIFNNLTSSPEGAVRLVTSQQASIALDRVPANVSKIAITVVIDGDTIS"
"
complement (6064..8154)
/gene="Z1617"
complement (6064..8154)
/gene="Z1617"
/function="putative membrane; Cell envelop: Outer membrane
constituents"
/notes="Residues 1 to 696 of 696 are 99.85 pct identical to
residues 1 to 696 of 696 from GenPept 118 :
gi|7108480|gb|AAF36432.1|AF126104_1 (AF126104) adhesin
[Escherichia coli O157:H7]"
/codon_start=1
/transl_table=11
/product="putative receptor"
/protein_id="AAG55732.1"

```

Query Match 99.9%; Score 2089.4; DB 1; Length 10029;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2090; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 ATGCGAATAACCACTCTGGCTTCGGTAGTCAATCCCTGTCTCGGATTTTCAGCCAGCAG 60
Db ATGCGAATAACCACTCTGGCTTCGGTAGTCAATCCCTGTCTCGGATTTTCAGCCAGCAG 8095
QY 61 ATAGCTCTCTCAGAGGATGTGATGATTTCTCGGCATCCGGCTATCAGAAAAGCTGACT 120
Db ATAGCTCTCTCAGAGGATGTGATGATTTCTCGGCATCCGGCTATCAGAAAAGCTGACT 8035
QY 121 AACCGAGCCGCCAGTGTCTGTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCAAC 180
Db AACCGAGCCGCCAGTGTCTGTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCAAC 7975
QY 181 GATCTGCGCGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGGTACGGGTAA 240
Db GATCTGCGCGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGGTACGGGTAA 7915
QY 241 ACCGAGGGCTGGAATTCAGCATCCGAGGAATSCCAGCCAGTATACACGCTGATCTGATT 300
Db ACCGAGGGCTGGAATTCAGCATCCGAGGAATSCCAGCCAGTATACACGCTGATCTGATT 7855
QY 301 GATGGTCTTCGTACAGGCGGAGACAGTACGTGACGTCCCAACGGTTTTCTCCCATGAT 360
Db GATGGTCTTCGTACAGGCGGAGACAGTACGTGACGTCCCAACGGTTTTCTCCCATGAT 7795
QY 361 ACCGGTTCATGCCCTCTGCGCCCATTCGAGCGTATTGAGGTTATCAGGGGGCGGATG 420
Db ACCGGTTCATGCCCTCTGCGCCCATTCGAGCGTATTGAGGTTATCAGGGGGCGGATG 7735
QY 421 TCACACTGTATGCGTCTGATCGATGGGCGGTGTGGTGAATATCATATACCAAGAGAT 480
Db TCACACTGTATGCGTCTGATCGATGGGCGGTGTGGTGAATATCATATACCAAGAGAT 7675
QY 481 GCAGACAAATGGCTCTCTTCCTCGTCAATCAGAGGCTGAAATCTCAGGAAAGCAAAATGG 540
Db GCAGACAAATGGCTCTCTTCCTCGTCAATCAGAGGCTGAAATCTCAGGAAAGCAAAATGG 7615
QY 541 GGTAAACAGCAGCCAGTTAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTCTGTCTCAG 600
Db GGTAAACAGCAGCCAGTTAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTCTGTCTCAG 7555
QY 601 CTCAGCTACGCGGTAGCACACAAACAGCGTCAAGGTTTCATCGGTTCACATCATCTGAGCGAT 660
Db CTCAGCTACGCGGTAGCACACAAACAGCGTCAAGGTTTCATCGGTTCACATCATCTGAGCGAT 7495
QY 661 ACACGAGGACCGGTATTCCTTATCCACGAGGATCAGAAATATAATCTTGGTGCAGCT 720
Db ACACGAGGACCGGTATTCCTTATCCACGAGGATCAGAAATATAATCTTGGTGCAGCT 7435
QY 721 CTTGACTGGAAGCGCTCGGAGCAGGATGTCCTGTGTTGATATGATACCAACCCGGCAG 780
Db CTTGACTGGAAGCGCTCGGAGCAGGATGTCCTGTGTTGATATGATACCAACCCGGCAG 7375
QY 781 CGTTATGATAACCGGATGGGCAACTGGGGAGTCTGACGGGGGATATGACCGGACCCCTG 840
Db CGTTATGATAACCGGATGGGCAACTGGGGAGTCTGACGGGGGATATGACCGGACCCCTG 7315
QY 841 CGCTATGAGCGAAACAAAATTTTCAGTGGCTATGATCATCTTTCACCTTCGGAACATGG 900
Db CGCTATGAGCGAAACAAAATTTTCAGTGGCTATGATCATCTTTCACCTTCGGAACATGG 7255
QY 901 AATTCGTATCTGAACTGGACAGCAGAGCAAAAATAAAGTCTGTGAGCTTGTACGAGTGTGA 960
Db AATTCGTATCTGAACTGGACAGCAGAGCAAAAATAAAGTCTGTGAGCTTGTACGAGTGTGA 7195
QY 961 CTGAAGCGGACAAAATGGGGGCTTTCGGGTACGCGCGGAGCTTAAGGAATCGAACCTT 1020
Db CTGAAGCGGACAAAATGGGGGCTTTCGGGTACGCGCGGAGCTTAAGGAATCGAACCTT 7135
QY 1021 ATCTGAATTCATTCTGCTTACCCGCTTGGGAGAAATCTCATCTGCTTACGTTGGGGGGC 1080
Db ATCTGAATTCATTCTGCTTACCCGCTTGGGAGAAATCTCATCTGCTTACGTTGGGGGGC 7075
QY 1081 GAGTTTCAGAGCTCGTCCATGAAACGAGGTGTGCTTTCGCCAGCAGCAGGTGAAACTTTC 1140

```


Db 7074 GAGTTTCAGAGCTCGTCCATGAAGAGCGAGTTCTCTTCCACACAGTGGAACATTTTC 7015

QY 1141 CGGCAGAAAAGCTGGTTCGTTATTTCTGAGATGAGTGGCATCTCAGGATGCACTTGGC 1200

Db 7014 CGGCAGAAAAGCTGGTTCGTTATTTCTGAGATGAGTGGCATCTCAGGATGCACTTGGC 6955

QY 1201 CTGACTCGGCGAGCGCTATGAACATCATGAGCAATTCGGGGGACATCTCAGTCCGCGT 1260

Db 6954 CTGACTCGGCGAGCGCTATGAACATCATGAGCAATTCGGGGGACATCTCAGTCCGCGT 6995

QY 1261 GCATATCTGGTCTGGGATGTGGCAGATGCTTGAAGCTGAAAGCGGTGTGACACAGGGA 1320

Db 6894 GCATATCTGGTCTGGGATGTGGCAGATGCTTGAAGCTGAAAGCGGTGTGACACAGGGA 6935

QY 1321 TATAAGGCACCGAATGGGCGACCTACATAAAGGGAATTAGTGGTGTGTCGGGCGAGGA 1380

Db 6834 TATAAGGCACCGAATGGGCGACCTACATAAAGGGAATTAGTGGTGTGTCGGGCGAGGA 6775

QY 1381 AAAACAATCTACTTGTAAACCCGACCTGAAGCCGAGAGAGCGTCAAGTTATGAGCT 1440

Db 6774 AAAACAATCTACTTGTAAACCCGACCTGAAGCCGAGAGAGCGTCAAGTTATGAGCT 6715

QY 1441 GGGGTGTATTACGATAACCCCGCGGTCTGAATGCCAATGTCAAGTGTATGACTGAC 1500

Db 6714 GGGGTGTATTACGATAACCCCGCGGTCTGAATGCCAATGTCAAGTGTATGACTGAC 6655

QY 1501 TTCTCCACAGATGTCTTATTCATTAATGAATGAACCAATAGCTATGTAACAGC 1560

Db 6654 TTCTCCACAGATGTCTTATTCATTAATGAATGAACCAATAGCTATGTAACAGC 6595

QY 1561 GGAAGGCCCGGTGTCACGGTGTGAATTTCCCGGACATTTGCGGTGTTGTCAGAGAT 1620

Db 6594 GGAAGGCCCGGTGTCACGGTGTGAATTTCCCGGACATTTGCGGTGTTGTCAGAGAT 6535

QY 1621 GTACCGTGTCACTGAATTAACCTGGAACCGGAAGTGAACAACTGATGTTGATACAA 1680

Db 6534 GTACCGTGTCACTGAATTAACCTGGAACCGGAAGTGAACAACTGATGTTGATACAA 6475

QY 1681 GGTGCGCGCTGAGTTATACCTTGAACACATGTTGAATGCGAACTGGAACCTGCGAGATC 1740

Db 6474 GGTGCGCGCTGAGTTATACCTTGAACACATGTTGAATGCGAACTGGAACCTGCGAGATC 6415

QY 1741 ACCGAAGATGGCATCAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1800

Db 6414 ACCGAAGATGGCATCAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 6355

QY 1801 CAGAATTTATTCGTCACCTGAGCGCTGTACAGAGAAAGTGTATGATGAGAGAGAGATAC 1860

Db 6354 CAGAATTTATTCGTCACCTGAGCGCTGTACAGAGAAAGTGTATGATGAGAGAGAGATAC 6295

QY 1861 CTGAAGCCCTGGACGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1920

Db 6294 CTGAAGCCCTGGACGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 6235

QY 1921 CTGAATCTCGGTGATTAACCTGCTCAACAGGATTAAGTGAAGTGAAGTGAAGTGAAGTGA 1980

Db 6234 CTGAATCTCGGTGATTAACCTGCTCAACAGGATTAAGTGAAGTGAAGTGAAGTGAAGTGA 6175

QY 1981 GCCGGTAAGAGTACGCTGTATGCGGTGATTTACTTCCAGACGGGATCATCAACACAGGA 2040

Db 6174 GCCGGTAAGAGTACGCTGTATGCGGTGATTTACTTCCAGACGGGATCATCAACACAGGA 6115

QY 2041 TATGTATACCTGAGCGAAATTAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2091

Db 6114 TATGTATACCTGAGCGAAATTAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 6064

RESULT 5
AE005274/c
LOCUS
DEFINITION
Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 98
of 155.

ACCESSION AE005274 AE005174
VERSION AE005274.1 GI:12513984
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
21074935
PUBMED 11206551
REFERENCE
2 (bases 1 to 10155)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1. 10155
/organism="Escherichia coli O157:H7 EDL933"
/mol_type="genomic DNA"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
misc_feature
<1..>10155
/note="O-island #43; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"
242..550
/gene="terF"
/note="synonym: Z1177"
242..550
/gene="terF"
/functions="orf; Adaptations, atypical conditions"
/note="Residues 1 to 102 of 102 are 100.00 pct identical to residues 1 to 102 of 102 from GenPept 118;
gi|710848|gb|AAF36433.1|AF126104_2 (AF126104) putative tellurium resistance protein A [Escherichia coli O157:H7]"
/codon_start=1
/transl_table=11
/product="partial putative phage inhibition, colicin resistance and tellurite resistance protein"
/protein_id="AAG55322.1"
/db_xref="GI:12513985"
/translation="MNLQSGNIPLOQSTIRLNLOYPKSGFKGEPDTCFLFMLNAQGK VSGSDSDFIFYNLSPEGAVRLVTGQQASIEIALDRVPANVSKIAITVVIGDETIS"
complement(1142..3232)
/genes="Z1178"
complement(1142..3232)
/genes="Z1178"
/function="putative membrane; Cell envelop: Outer membrane constituents"
/note="Residues 1 to 696 of 696 are 99.85 pct identical to residues 1 to 696 of 696 from GenPept 118;
gi|7108480|gb|AAF36432.1|AF126104_1 (AF126104) adhesin [Escherichia coli O157:H7]"
/codon_start=1
/transl_table=11
/product="putative receptor"
/protein_id="AAG55323.1"
/db_xref="GI:12513986"

gene
CDS


```

/translation="MRITTLASVVIPLGFSASSIAAEDMIVSASGYEKKLTNAA
SVVISQELQSSYDLAALASVEGVDSVSGTGGTGLEISIRGMPASVYLLIIDG
VROGSSDVPTNGSANTGMPPLAIEIRIVIRGPMSTLYGSDAMGQVNIITRKN
ADKLLSVNAGLNQENKNGNSQNFWSGSLVDVDSVLSQVRGSDQORQSSVTSL
SDTAATPIPTESQSNVNLGARLDKASEQDLVDFMDTTRQRYDNRDQSLGTVGY
LTERLYERNKISAGYDHTFTFTWKSJWNETENGRELVRSYLRDKNGLAGPRE
LKEENLLINSLIPLGESHVTVGSEFQSSNMKGVLASTGEIFRQKSWSVFAEDE
WHLDALATAGRIEHEQFGHFSPRAYLVWDVADAWTLKGGVTGTYKAPRMQQLH
KISGVSQQRKLLGNPDLPKPEESVYEAGVYDNPAGLNANVTGFTDFKNIQVY
SINDTNSVNSKARLHGVEFAGTLPNSVEDTSLNVTWTRSEQRDGNKGLPSY
TPEHVNAKLNQITTEEVASWLGARYGKTPRTQNYSSLSAVQKVKYDEKEVYLKAW
TVVDGLSWKMTDALTLNAAVNNLNKDYSDVLSYAGKSTLYAGDYFQTGSSITGVY
IPERNYWSLNYQF"
/ gene 3367 3429
/ gene="Z1179"
/ gene="Z1179"
/ function="orf; Unknown function"
/ note="No significant matches"
/ codon_start=1
/ transl_table=11
/ protein_id="AAG55324.1"
/ db_xref="GI:12513987"
/ translation="MPVNSYPYGVFGFANKKWL"
/ complement(3902..4189)
/ gene="Z1180"
/ complement(3902..4189)
/ function="orf; Unknown function"
/ note="No significant matches"
/ codon_start=1
/ transl_table=11
/ protein_id="AAG55325.1"
/ db_xref="GI:12513988"
/ translation="MSVINDRMSGKTPFVDGNLFLRPHRYMRDFFFOTHPEMG
QAWFOEIRNTLGYCSLKSILICNSNPEPVKRGVTSILKIKVVMCIKPEGL"
/ 4686 4904
/ gene="Z1181"
/ 4686..4904
/ gene="Z1181"
/ function="orf; Unknown function"
/ notes="Residues 1 to 58 of 72 are 41.37 pct identical to
residues 159 to 216 of 229 from GenPept 118 :
gi|5880473|gb|AAD54666.1|AF097520_4 (AF097520) unknown
[Shigella flexneri]"
/ codon_start=1
/ transl_table=11
/ protein_id="AAG55326.1"
/ db_xref="GI:12513989"
/ translations="MPKYAEMLMAEEDDKVCBLIFKQDNLLIKINVRPPIIKKYR
TLOCKVVMWILPFIWELIVCTLVGNS"
/ complement(5306..5404)
/ gene="Z1182"
/ complement(5306..5404)
/ gene="Z1182"
/ function="orf; Unknown function"
/ note="No significant matches"
/ codon_start=1
/ transl_table=11
/ protein_id="AAG55327.1"
/ db_xref="GI:12513990"
/ translations="MIIITLIGIHILQRIYINWYSGNYILKIQ"
/ 5537..5872
/ gene="Z1183"
/ 5537..5872
/ gene="Z1183"
/ function="orf; Unknown function"
/ notes="Residues 68 to 111 of 111 are 50.00 pct identical
to residues 21 to 54 of 90 from GenPept 118 :
gi|3661496|gb|AAC61726.1| (AF081285) neurotensin receptor
R8 [Escherichia coli]"
/ codon_start=1
/ transl_table=11

```

```

/ protein_id="AAG55328.1"
/ db_xref="GI:12513991"
/ translations="MVISVGYEKRTGSSASSIIIQKLOPDNAILIQKKAISVGRN
VEIATKRGSDINTRIPASNCLSARNALSKHNHTGARILMSYAFSIVITNENKLAG
ANIKRCDTV"
/ 6389..6523
/ gene="Z1184"
/ 6389..6523
/ gene="Z1184"
/ function="orf; Unknown function"
/ note="No significant matches"
/ codon_start=1
/ transl_table=11
/ protein_id="AAG55329.1"
/ db_xref="GI:12513992"
/ translation="MLQIMTRWISLDAYDYMHHVNDYATECMTHIKALGSLITWSC"
/ complement(6653..6847)
/ gene="Z1185"
/ complement(6653..6847)
/ gene="Z1185"
/ function="orf; Unknown function"
/ note="No significant matches"
/ codon_start=1
/ transl_table=11
/ protein_id="AAG55330.1"
/ db_xref="GI:12513993"
/ translation="MKNKIKRLQLQARQREFATKNGRAALPHERLTYFPFLEDLIK
NNGKKIRSKKHFLKNEBK"
/ 7167..7439
/ gene="Z1186"
/ 7167..7439
/ gene="Z1186"
/ function="orf; Unknown function"
/ note="No significant matches"
/ codon_start=1
/ transl_table=11
/ protein_id="AAG55331.1"
/ db_xref="GI:12513994"
/ translation="MSYQPGVLPLDLRRGYLSKNASSNRFGLVNGDGLSKDRCW
CTPSRWRSSMDNAIACSSSLIVCTWEALYRSTSHYGNKXNLLQLA"
/ 7831..7938
/ gene="Z1187"
/ 7831..7938
/ gene="Z1187"
/ function="orf; Unknown function"
/ note="No significant matches"
/ codon_start=1
/ transl_table=11
/ protein_id="AAG55332.1"
/ db_xref="GI:12513995"

```

```

Query Match 99.9%; Score 2089.4; DB 1; Length 10155;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2090; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATCGAATAACCACTCTGGCTTCGCTAGTCATTCCCTGTCTCGGATTTTCACCCAGCAGC 60
Db 3232 ATCGAATAACCACTCTGGCTTCGCTAGTCATTCCCTGTCTCGGATTTTCACCCAGCAGC 3173

Qy 61 ATAGCTCTGCAGAGGATGTGATGTCTCGGCATCCGGCTATGAGAAAAGCTGACT 120
Db 3172 ATAGCTCTGCAGAGGATGTGATGTCTCGGCATCCGGCTATGAGAAAAGCTGACT 3113

Qy 121 AACGCAAGCCAGTCTTTCTGTGATTACGAGGAGGAATTCAGTCCAGCCAGTACCAC 180
Db 3112 AACGCAAGCCAGTCTTTCTGTGATTACGAGGAGGAATTCAGTCCAGCCAGTACCAC 3053

Qy 181 GATCTCGCGGAGGCTCTGAGATCAGTAGAGGTGTGGATGTTGAAAGTGGTACGGGTAAA 240
Db 3052 GATCTCGCGGAGGCTCTGAGATCAGTAGAGGTGTGGATGTTGAAAGTGGTACGGGTAAA 2993

Qy 241 ACCGGAGGCGTGTGAATCAGCATCCGAGGATCCGAGGAGTTACACGCTGATCTGATT 300

```


Db 2992 ACCGAGGGCTGGAATCAGATCCGAGGAATCCAGCCAGTTACACGCTGATCTGATT 2933
Qy 301 GATGGTGTTCGTAGGGCGGAGCAGTGAAGTGAATCCCAACGGTTTTTTCGCCATGAAT 360
Db 2932 GATGGTGTTCGTAGGGCGGAGCAGTGAAGTGAATCCCAACGGTTTTTTCGCCATGAAT 2873
Qy 361 ACCGGGTTTCATGCCCCCTCTGCGCGCCATTGAGCGTATTCAGGTTATCAGGGGCGCATG 420
Db 2872 ACCGGGTTTCATGCCCCCTCTGCGCGCCATTGAGCGTATTCAGGTTATCAGGGGCGCATG 2813
Qy 421 TCCACATGTATGCTCTGATGCGATGCGCGGTGTGGTGAATATCAATACAGAAAGAAT 480
Db 2812 TCCACATGTATGCTCTGATGCGATGCGCGGTGTGGTGAATATCAATACAGAAAGAAT 2753
Qy 481 GCACACAAATGGCTCTCTCCGTCATGACGGGCTGNAATCTGCAGGAAGCAACAATGG 540
Db 2752 GCACACAAATGGCTCTCTCCGTCATGACGGGCTGNAATCTGCAGGAAGCAACAATGG 2693
Qy 541 GGTAAACAGCAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCCTTGTGGATGATCTGTCAAGC 600
Db 2692 GGTAAACAGCAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCCTTGTGGATGATCTGTCAAGC 2633
Qy 601 CTGACAGTACGGGTAGCAGACACACAGCGTCAAGGTTTCATCGGTCAATCATCTAGCGAT 660
Db 2632 CTGACAGTACGGGTAGCAGACACACAGCGTCAAGGTTTCATCGGTCAATCATCTAGCGAT 2573
Qy 661 ACAGCAGGACCGGTATTTCTTATCCACGAGTTCACAGAAATATATCTTGTGTCAAGT 720
Db 2572 ACAGCAGGACCGGTATTTCTTATCCACGAGTTCACAGAAATATATCTTGTGTCAAGT 2513
Qy 721 CTTGACAGGAGGCTCGGAGCAGGATGTCTCTGTTTATATGATACCAACCCGGCAG 780
Db 2512 CTTGACAGGAGGCTCGGAGCAGGATGTCTCTGTTTATATGATACCAACCCGGCAG 2453
Qy 781 CGTTATGATACCGGAGTGGCACTGGGAGTCTGACGGGGGATGATGACGGACCCCTG 840
Db 2452 CGTTATGATACCGGAGTGGCACTGGGAGTCTGACGGGGGATGATGACGGACCCCTG 2393
Qy 841 CGTTATGAGCAAAACAAAATTTTCAGCTGGCTATGATCATATCTTCACTTCGGAACATGG 900
Db 2392 CGTTATGAGCAAAACAAAATTTTCAGCTGGCTATGATCATATCTTCACTTCGGAACATGG 2333
Qy 901 AATTCGTATCTGAATCTGACGACGACAGCAAGAAATTAAGGTCTGAGCTTGTACGAGTGA 960
Db 2332 AATTCGTATCTGAATCTGACGACGACAGCAAGAAATTAAGGTCTGAGCTTGTACGAGTGA 2273
Qy 961 CTGAAGCGCACAAATGGGGGCTTGGCGGTGACCGCGGAGCTTAAGGAATCGAACCTT 1020
Db 2272 CTGAAGCGCACAAATGGGGGCTTGGCGGTGACCGCGGAGCTTAAGGAATCGAACCTT 2213
Qy 1021 ATCCTGAATTCATCTGCTTACCCCTCTGGGAGAAATCTCATCTGCTGATACGGTGGGGGC 1080
Db 2212 ATCCTGAATTCATCTGCTTACCCCTCTGGGAGAAATCTCATCTGCTGATACGGTGGGGGC 2153
Qy 1081 GAGTTTCAGAGCTCGCTCAAGAGACGAGTGTCTCTGCGCAGACAGCTGAACCTTTC 1140
Db 2152 GAGTTTCAGAGCTCGCTCAAGAGACGAGTGTCTCTGCGCAGACAGCTGAACCTTTC 2093
Qy 1141 CGGCAAGAAAGCTGTCGCTGATTTGCTGAGGATGAGTGGCATCTCACGGATGCACTTCG 1200
Db 2092 CGGCAAGAAAGCTGTCGCTGATTTGCTGAGGATGAGTGGCATCTCACGGATGCACTTCG 2033
Qy 1201 CTGACTCGGGCAGCGCTATGACATCATGACCAATTCGGGGACCTTCAGTCCGGT 1260
Db 2032 CTGACTCGGGCAGCGCTATGACATCATGACCAATTCGGGGACCTTCAGTCCGGT 1973
Qy 1261 GCATATCTGCTGCGATGCGCATGCTTGGACGCTGAAAGCGGCTGTGACCAACGGGA 1320
Db 1972 GCATATCTGCTGCGATGCGCATGCTTGGACGCTGAAAGCGGCTGTGACCAACGGGA 1913
Qy 1321 TATAGGCACCGATGCGGAGCTTACATAAGGATGATGCTGCTGCTGCGGAGGGA 1380

Db 1912 TATAAGGCACCCAGCAATGGGGCAGCTACTATAAGGATTAAGTGTGCTGCTGGGAGGGA 1853
Qy 1381 AAAACAAATCTACTTGGTAACCCCGACCTGAAGCGGAGAGAGCGTCACTTATGAGCT 1440
Db 1852 AAAACAAATCTACTTGGTAACCCCGACCTGAAGCGGAGAGAGCGTCACTTATGAGCT 1793
Qy 1441 GGGGTGATTAAGATAACCCCGCGCTGTAATCCCAATGTCAAGGTTTATGACTGAC 1500
Db 1792 GGGGTGATTAAGATAACCCCGCGCTGTAATCCCAATGTCAAGGTTTATGACTGAC 1733
Qy 1501 TTCTCCAAACAAGATGTCTCTTATTCATAAATATAACCAATAGCTATGTAACAGC 1560
Db 1732 TTCTCCAAACAAGATGTCTCTTATTCATAAATATAACCAATAGCTATGTAACAGC 1673
Qy 1561 GGAAAGGCCGGTTGCACGGTGTGGAATTTGCGGCGACATTCGCGCTGTGTGCTAGAGAT 1620
Db 1672 GGAAAGGCCGGTTGCACGGTGTGGAATTTGCGGCGACATTCGCGCTGTGTGCTAGAGAT 1613
Qy 1621 GTCACGCTGTCACTGAATTAACCTGGAACCCGAAAGTGAACAACTGATGTTGTAACAAA 1680
Db 1612 GTCACGCTGTCACTGAATTAACCTGGAACCCGAAAGTGAACAACTGATGTTGTAACAAA 1553
Qy 1681 GGTGCGCGCTGAGTATTAACCCCTGGAACACATGTGTAATGGAACCTGAACCTGACAGATC 1740
Db 1552 GGTGCGCGCTGAGTATTAACCCCTGGAACACATGTGTAATGGAACCTGAACCTGACAGATC 1493
Qy 1741 ACCGAAGAGGTGGCATCATGCTGGTGGCGCTTATCGCGGAAACACACAGCTTTCACC 1800
Db 1492 ACCGAAGAGGTGGCATCATGCTGGTGGCGCTTATCGCGGAAACACACAGCTTTCACC 1433
Qy 1801 CAGAATTAATTCGTCACTGAGCGCTGTACAGAAAGAGTGTATGATGAGAAGGAGATAC 1860
Db 1432 CAGAATTAATTCGTCACTGAGCGCTGTACAGAAAGAGTGTATGATGAGAAGGAGATAC 1373
Qy 1861 CTGAAGCCTGGACGGTGGTGGATGACAGTCTGCTGGAAGATGACGGATGCTGACG 1920
Db 1372 CTGAAGCCTGGACGGTGGTGGATGACAGTCTGCTGGAAGATGACGGATGCTGACG 1313
Qy 1921 CTGAATCTGCGGTGAATAACCTGCTCAACAAGATTTACAGTACGCTGACCTGTACAGT 1980
Db 1312 CTGAATCTGCGGTGAATAACCTGCTCAACAAGATTTACAGTACGCTGACCTGTACAGT 1253
Qy 1981 GCCGTAAGAGTACCTGTATGCGGCTGATTAATCCAGAGATTTACAGTACGCTGACAGG 2040
Db 1252 GCCGTAAGAGTACCTGTATGCGGCTGATTAATCCAGAGATTTACAGTACGCTGACAGG 1193
Qy 2041 TATGTGATCTGACGCGAAATTAATGATGCTGCTGAACCTATCAGTTCCTGA 2091
Db 1192 TATGTGATCTGACGCGAAATTAATGATGCTGCTGAACCTATCAGTTCCTGA 1142

RESULT 6
AR204161/cLOCUS AR204161 87563 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 57 from patent US 6365723.

ACCESSION AR204161

VERSION AR204161.1 GI:21500738

KEYWORDS

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 87563)

AUTHORS Blattner, F.R., Buxland, V., Perna, N.T., Plunkett, G. and Welch, R.

TITLE Sequences of E. coli O157

JOURNAL Patent: US 6365723-A 57 02-APR-2002;

FEATURES Location/Qualifiers

source 1..87563

/organism="unknown"

/mol_type="unassigned DNA"

ORIGIN

Query Match 99.9%; Score 2089.4; DB 6; Length 87563;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 2090; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Qy	1 ATGCGAATAACCACTCTGGCTTCCTAGTCAATCCCTCTCGGATTTTCAGCAGCAGC 60
Dd	49991 ATGCGAATAACCACTCTGGCTTCCTAGTCAATCCCTCTCGGATTTTCAGCAGCAGC 49932
Qy	61 ATAGTCTGCTGAGAGATGTGATGATCTCTCGGCATCCGCTATGAGAAAGCTGACT 120
Dd	49931 ATAGTCTGCTGAGAGATGTGATGATCTCTCGGCATCCGCTATGAGAAAGCTGACT 49872
Qy	121 AACGAGCGCCGAGTGTCTCTGATTAGCCAGAGGAATTGCACTCCAGCCAGTACCAC 180
Dd	49871 AACGAGCGCCGAGTGTCTCTGATTAGCCAGAGGAATTGCACTCCAGCCAGTACCAC 49812
Qy	181 GATCTGCGGAGGCTCTGAGATCAGTACAGGCTGTGATGTTGAAAGTGTACCGGTAAA 240
Dd	49811 GATCTGCGGAGGCTCTGAGATCAGTACAGGCTGTGATGTTGAAAGTGTACCGGTAAA 49752
Qy	241 ACCGAGGCTCGAATCAGCATCCGAGGAATGCCAGGCAATTCACGCTGATCTGATT 300
Dd	49751 ACCGAGGCTCGAATCAGCATCCGAGGAATGCCAGGCAATTCACGCTGATCTGATT 49692
Qy	301 GATGCTGCTGAGCGGGAAGCTGACGTGACTCCCAACGGTTTTTCGCCATGAAT 360
Dd	49691 GATGCTGCTGAGCGGGAAGCTGACGTGACTCCCAACGGTTTTTCGCCATGAAT 49632
Qy	361 ACCGGTTCATGCCCTCTGCGCCATTGAGCGTATTGAGGTTATCAGGGGCGCATG 420
Dd	49631 ACCGGTTCATGCCCTCTGCGCCATTGAGCGTATTGAGGTTATCAGGGGCGCATG 49572
Qy	421 TCCACACTGATGGCTCTGATCGATGGCGGTGTGGTGAATATCATTTACGAAAGAAAT 480
Dd	49571 TCCACACTGATGGCTCTGATCGATGGCGGTGTGGTGAATATCATTTACGAAAGAAAT 49512
Qy	481 GCGAATAATGCTCTCTTCGTCATATGACGGCTGAATCTCGAGGAAGCAAAATGG 540
Dd	49511 GCGAATAATGCTCTCTTCGTCATATGACGGCTGAATCTCGAGGAAGCAAAATGG 49452
Qy	541 GGTAAACAGCAGCAGTAAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTTCTGTACG 600
Dd	49451 GGTAAACAGCAGCAGTAAATTTCTGGAGCAGTGGTCCCTTTGTGGATGATTTCTGTACG 49392
Qy	601 CTCAGGATCGCGGTAGCACAACAGCTCAGGTTTCATCGGTACATCACTGACGAT 660
Dd	49391 CTCAGGATCGCGGTAGCACAACAGCTCAGGTTTCATCGGTACATCACTGACGAT 49332
Qy	661 ACAGCGGACGCGTATTCCTTATCCACGAGTCAAGATTTATATTTCTGGTGACGT 720
Dd	49331 ACAGCGGACGCGTATTCCTTATCCACGAGTCAAGATTTATATTTCTGGTGACGT 49272
Qy	721 CTTGACTGGAAGCGTCGGAGCAGATGTCTCTGGTTGATATGATACCCCGGAG 780
Dd	49271 CTTGACTGGAAGCGTCGGAGCAGATGTCTCTGGTTGATATGATACCCCGGAG 49212
Qy	781 GGTATGATTAACCGGATGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCTG 840
Dd	49211 GGTATGATTAACCGGATGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCTG 49152
Qy	841 CGCTATCAGCGAACAATAATTTCTGCTGGCTATGATCATCTTTCCCTTCGGAACATGG 900
Dd	49151 CGCTATCAGCGAACAATAATTTCTGCTGGCTATGATCATCTTTCCCTTCGGAACATGG 49092
Qy	901 AAATCGTATCTGAATCGAACGAGACAGAAATAAGGCTGTGAGCTTTGACCGAGTGA 960
Dd	49091 AAATCGTATCTGAATCGAACGAGACAGAAATAAGGCTGTGAGCTTTGACCGAGTGA 49032
Qy	961 CTGAAGCGCACAATAATGGGGCTTCCGCTCAGCGCGGAGCTTAAGGAATCGAACCTT 1020
Dd	49031 CTGAAGCGCACAATAATGGGGCTTCCGCTCAGCGCGGAGCTTAAGGAATCGAACCTT 48972
Qy	1021 ATCCTGAATTCATTACTGCTTACCCCTCTGGGAGAAATCTCATCTGGTTACGTTGGGGGC 1080
Dd	48971 ATCCTGAATTCATTACTGCTTACCCCTCTGGGAGAAATCTCATCTGGTTACGTTGGGGGC 48912

RESULT 7

BD184766/c

LOCUS

86248 bp

DNA

linear

PAT 17-JUN-2003

Qy	1081 GAGTTTCAGAGCTCGTCCATGAAAGACGAGATGTCTCTGCCAGCACAGGTGAACTTTC 1140
Dd	48911 GAGTTTCAGAGCTCGTCCATGAAAGACGAGATGTCTCTGCCAGCACAGGTGAACTTTC 48852
Qy	1141 CGCAGAAAAGCTGGTCCGATTTTGTGTAGGATGAGTGGCACTCTCAGGATGCACTTGG 1200
Dd	48851 CGCAGAAAAGCTGGTCCGATTTTGTGTAGGATGAGTGGCACTCTCAGGATGCACTTGG 48792
Qy	1201 CTGACTGCGGGCAGCGCTATGAAATCATGAGCAATTCGGGGGACACTTCAGTCCGCGT 1260
Dd	48791 CTGACTGCGGGCAGCGCTATGAAATCATGAGCAATTCGGGGGACACTTCAGTCCGCGT 48732
Qy	1261 GCATATCTGCTCTGGGATGTCGAGATGCTGTGAGCGTGAAGCGCGGTGTGACACGGGA 1320
Dd	48731 GCATATCTGCTCTGGGATGTCGAGATGCTGTGAGCGTGAAGCGCGGTGTGACACGGGA 48672
Qy	1321 TATAAGCACCCCAAGATGGGGCAGCTACATAAAGGATTAAGTGTGTCCGGGGCAGGGA 1380
Dd	48671 TATAAGCACCCCAAGATGGGGCAGCTACATAAAGGATTAAGTGTGTGTCCGGGGCAGGGA 48612
Qy	1381 AAAACAAATCTACTTGGTAAACCCCGACCTGAAGCCGGAAGAGAGCGTCACTTATGAGGCT 1440
Dd	48611 AAAACAAATCTACTTGGTAAACCCCGACCTGAAGCCGGAAGAGAGCGTCACTTATGAGGCT 48552
Qy	1441 GGGGTGTATTACGATTAACCCCGCGCTCTGAATGCCAAATGTCACAGTGTATGACTGAC 1500
Dd	48551 GGGGTGTATTACGATTAACCCCGCGCTCTGAATGCCAAATGTCACAGTGTATGACTGAC 48492
Qy	1501 TTCTCCAAACAAAGATTGCTCTTATCCATAAATGATAACCAATAGCTATGTAACAGC 1560
Dd	48491 TTCTCCAAACAAAGATTGCTCTTATCCATAAATGATAACCAATAGCTATGTAACAGC 48432
Qy	1561 GGAAAGCCCGGTTGCAAGTGTGGAAATTTCCGGGCACTTCCGCTGTGGTCAAGGAT 1620
Dd	48431 GGAAAGCCCGGTTGCAAGTGTGGAAATTTCCGGGCACTTCCGCTGTGGTCAAGGAT 48372
Qy	1621 GTCAAGCTGTCACTGAATTAACCTGACCCGGAAGTGAACAACTGATGTGTGATAACAAA 1680
Dd	48371 GTCAAGCTGTCACTGAATTAACCTGACCCGGAAGTGAACAACTGATGTGTGATAACAAA 48312
Qy	1681 GGTGCGCGCTGAGTTATACCCCTGAACACATGATGTAATGCAAACTGAATGCGCAGATC 1740
Dd	48311 GGTGCGCGCTGAGTTATACCCCTGAACACATGATGTAATGCAAACTGAATGCGCAGATC 48252
Qy	1741 ACCGAAGAGTGGCATCATGCTGGGTCCCGTTATCGCGGGAACACACAGCTTTTACC 1800
Dd	48251 ACCGAAGAGTGGCATCATGCTGGGTCCCGTTATCGCGGGAACACACAGCTTTTACC 48192
Qy	1801 CAGAAATATTCTGCTCACTGAGCGCTGTACAGAAAGAAAGTGTATGATGAGAAAGGAGATAC 1860
Dd	48191 CAGAAATATTCTGCTCACTGAGCGCTGTACAGAAAGAAAGTGTATGATGAGAAAGGAGATAC 48132
Qy	1861 CTGAAAGCTGGAAGGCTGGATGCGAGTCTGTCTGGAAGATGACGGATGCCCTGACG 1920
Dd	48131 CTGAAAGCTGGAAGGCTGGATGCGAGTCTGTCTGGAAGATGACGGATGCCCTGACG 48072
Qy	1921 CTGANTGCTGCGGTGAATAAATCTGCTCAACAGGATTAACAGTACGCTGAGCTGTACAGT 1980
Dd	48071 CTGANTGCTGCGGTGAATAAATCTGCTCAACAGGATTAACAGTACGCTGAGCTGTACAGT 48012
Qy	1981 GCGGTAAGAGTACGCTGTATTCGCGGTGATTACTTCCAGAGGGGATCATCAACACAGGA 2040
Dd	48011 GCGGTAAGAGTACGCTGTATTCGCGGTGATTACTTCCAGAGGGGATCATCAACACAGGA 47952
Qy	2041 TATGTATACCTGAGCGAAATTAAGTGTGCTGCTGAACTATCACTGTTCTGA 2091
Dd	47951 TATGTATACCTGAGCGAAATTAAGTGTGCTGCTGAACTATCACTGTTCTGA 47901


```

DEFINITION Nucleic acid molecule and polypeptide specific to intestinal
hemorrhagic pathogenic Escherichia coli O157:H7, and method of use
thereof.
ACCESSION BD184766
VERSION 1
KEYWORDS JP 2002355074-A/2,
SOURCE Escherichia coli O157:H7
ORGANISM Escherichia coli O157:H7
REFERENCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
AUTHORS Enterobacteriaceae; Escherichia.
1 (bases 1 to 86248)
Hayashi, H., Shinagawa, H., Makino, K., Hayashi, T., Onishi, S.,
Hattori, M. and Kurokawa, K.
TITLE Nucleic acid molecule and polypeptide specific to intestinal
hemorrhagic pathogenic Escherichia coli O157:H7, and method of use
JOURNAL Patent: JP 2002355074-A 2 10-DEC-2002;
COMMENT PRESIDENT OF UNIVERSITY OF TSUKUBA
OS Escherichia coli O157:H7
PN JP 2002355074-A/2
PD 10-DEC-2002
PF 24-JAN-2002 JP 2002015959
PI HIDEO HAYASHI, HIDEO SHINAGAWA, KOZO MAKINO, TETSUYA HAYASHI, SHIN
ONISHI,
PI MASAHIRA HATTORI, KEN KUROKAWA
PC C12N15/09, C12N15/09, A61K31/7088, A61K39/00, A61K48/00, A61P31/04,
PC C07K14/245,
PC C07K16/12, C12M1/00, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12P21/
PC 02, C12Q1/68,
PC G01N33/15, G01N33/50, G01N33/53, G01N33/53, G01N33/566, G01N37/00,
PC C12N15/00,
PC C12N15/00, C12N5/00
CC Nucleic acid molecule and polypeptide specific to intestinal
hemorrhagic
CC pathogenic Escherichia coli O157:H7, and method of use thereof
FH Key Location/Qualifiers
FT source 1..86248
FT /organism= 'Escherichia coli O157:H7'.
FT Location/Qualifiers
1..86248
/organism= 'Escherichia coli O157:H7'
/mol_type= 'genomic DNA'
/db_xref= 'taxon:83334'

Query Match 99.8%; Score 2087.8; DB 6; Length 86248;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2089; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCGGAATACCACTCTGGCTTCGGTAGTCATTCCTGCTCGGATTTTCAGCCAGCAGC 60
DB 50004 ATCGGAATACCACTCTGGCTTCGGTAGTCATTCCTGCTCGGATTTTCAGCCAGCAGC 49945

QY 61 ATAGCTGCTCAGAGGATGTGATGTCTCGGCATCCGGCTATGAGAAAAGCTGACT 120
DB 49944 ATAGCTGCTCAGAGGATGTGATGTCTCGGCATCCGGCTATGAGAAAAGCTGACT 49885

QY 121 AACGACGCCAGTGTCTGTGATAGCCAGGAGGAATTCAGTCCAGCCAGTACCAC 180
DB 49884 AACGACGCCAGTGTCTGTGATAGCCAGGAGGAATTCAGTCCAGCCAGTACCAC 49825

QY 181 GATCTCGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGTTACGGGTAAA 240
DB 49824 GATCTCGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGTTACGGGTAAA 49765

QY 241 ACCGAGGGCTGGAATCAGATCCGAGGAATGCCAGCCAGTTACACGCTGATCTGATT 300
DB 49764 ACCGAGGGCTGGAATCAGATCCGAGGAATGCCAGCCAGTTACACGCTGATCTGATT 49705

QY 301 GATGGTGTTCGTTCAGGGCGGAGCAGTACGCTGCCAACCGGTTTTCTGCCATGAAT 360
DB 49704 GATGGTGTTCGTTCAGGGCGGAGCAGTACGCTGCCAACCGGTTTTCTGCCATGAAT 49645

```


Db	48564	GGGGTATTACGATAACCCCGCGGCTCTAAATGCCAATGTACAGGTTTATGACTGAC	48505
Qy	1501	TTTCTCAACAAGATTGTCTTTATTCATAAATGATAACCAATAGCTATGTAACAGC	1560
Db	48504	TTTCTCAACAAGATTGTCTTTATTCATAAATGATAACCAATAGCTATGTAACAGC	48445
Qy	1561	GGAAGGCCCGGTTGCGAGTGTGGATTTGCGGACATTTGCGGCTGTGGTCAGAGAT	1620
Db	48444	GGAAGGCCCGGTTGCGAGTGTGGATTTGCGGACATTTGCGGCTGTGGTCAGAGAT	48385
Qy	1621	GTACGCTGTCTCACTGAATTAACCTGACCGGAGTGAACAAAGTGAACAAAGTGAACAA	1680
Db	48384	GTACGCTGTCTCACTGAATTAACCTGACCGGAGTGAACAAAGTGAACAAAGTGAACAA	48325
Qy	1681	GGTGCGCGCTGAGTTATACCCCTGAACATGATGTAATGGAATGGAATGGAATGGAATG	1740
Db	48324	GGTGCGCGCTGAGTTATACCCCTGAACATGATGTAATGGAATGGAATGGAATGGAATG	48265
Qy	1741	ACCAAGAGTGGCATCATGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG	1800
Db	48264	ACCAAGAGTGGCATCATGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG	48205
Qy	1801	CAGAATTATTCCTGACTGAGCGCTGTACAGAGAAAGTGTATGATGAGAAAGGAGATAC	1860
Db	48204	CAGAATTATTCCTGACTGAGCGCTGTACAGAGAAAGTGTATGATGAGAAAGGAGATAC	48145
Qy	1861	CTGAAGCCTGGACGCGGTGATACAGAGAAAGTGTATGATGAGAAAGGAGATAC	1920
Db	48144	CTGAAGCCTGGACGCGGTGATACAGAGAAAGTGTATGATGAGAAAGGAGATAC	48085
Qy	1921	CTGAATGCTCGGTTGATACCTGCTCAACAGAGTACAGTACAGTACAGTACAGTACAGT	1980
Db	48084	CTGAATGCTCGGTTGATACCTGCTCAACAGAGTACAGTACAGTACAGTACAGTACAGT	48025
Qy	1981	GCCGTTAAGAGTACGCTGTATGCGGTGATGCTCTGCGGAGATGACGATGACGATGACG	2040
Db	48024	GCCGTTAAGAGTACGCTGTATGCGGTGATGCTCTGCGGAGATGACGATGACGATGACG	47965
Qy	2041	TATGATACCTGAGCGAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT	2091
Db	47964	TATGATACCTGAGCGAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT	47914
RESULT 8			
AP002554/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			

Db 282695 CTGACGGTACGGGTAGCACACAACAGCGTCAAGGTTTCATCGGTCAATCACTAGACGAT 282636
QY 661 ACACGAGGACCGGTATTCCTTTATCCACGGAGTCAAGAAATTAATCTTGGTCCAGT 720
Db 282635 ACACGAGGACCGGTATTCCTTTATCCACGGAGTCAAGAAATTAATCTTGGTCCAGT 282576
QY 721 CTTGACTGAAGCGCTCGAGCAGAGATGCTCTGGTTTGTATGATGATACCAACCGGCGAG 780
Db 282575 CTTGACTGAAGCGCTCGAGCAGAGATGCTCTGGTTTGTATGATGATACCAACCGGCGAG 282516
QY 781 CTTTATGATAACCGGATGGGCAACTGGGGAGTCTGACGGGGGATATGACCGGACCCCTG 840
Db 282515 CTTTATGATAACCGGATGGGCAACTGGGGAGTCTGACGGGGGATATGACCGGACCCCTG 282456
QY 841 CGCTATGAGGAAACAAATTTTCAGCTGCTATGATCATCTTTTACCTTCGACATGG 900
Db 282455 CGCTATGAGGAAACAAATTTTCAGCTGCTATGATCATCTTTTACCTTCGACATGG 282396
QY 901 AAATCGTATCTGAACTGGAAACAGACAGAAATTAAGGTCGTGAGCTTGTACGAGTGTGA 960
Db 282395 AAATCGTATCTGAACTGGAAACAGACAGAAATTAAGGTCGTGAGCTTGTACGAGTGTGA 282336
QY 961 CTGAAGCGGACAAATGGGGGTTGCCGTGACCGCGGGAGCTTAAGGAATCGAACCTT 1020
Db 282335 CTGAAGCGGACAAATGGGGGTTGCCGTGACCGCGGGAGCTTAAGGAATCGAACCTT 282276
QY 1021 ATCCGAAATTCATTACTGCTTACCCCTCTGGGAGATCTCATCTGTTACGTTGGGGGC 1080
Db 282275 ATCCGAAATTCATTACTGCTTACCCCTCTGGGAGATCTCATCTGTTACGTTGGGGGC 282216
QY 1081 GAGTTTCAGAGCTCGTCCATGAAGACGGAGTTGCTTCCGACGACAGTGAAACTTTC 1140
Db 282215 GAGTTTCAGAGCTCGTCCATGAAGACGGAGTTGCTTCCGACGACAGTGAAACTTTC 282156
QY 1141 CGGCAAGAAAGCTGCTGCTGATTTTCTGAGGATGATGCGATCTCACGATCGACTGCG 1200
Db 282155 CGGCAAGAAAGCTGCTGCTGATTTTCTGAGGATGATGCGATCTCACGATCGACTGCG 282096
QY 1201 CTGACTCGGGGACCGCTATGAACATCATGACCAATTCGGGGGACACTTCAGTCCGGT 1260
Db 282095 CTGACTCGGGGACCGCTATGAACATCATGACCAATTCGGGGGACACTTCAGTCCGGT 282036
QY 1261 GCATATCTGCTGCGGATGTGGCAGATGCTGAGACGCTGAAAGCGGTTGACACGGGA 1320
Db 282035 GCATATCTGCTGCGGATGTGGCAGATGCTGAGACGCTGAAAGCGGTTGACACGGGA 281976
QY 1321 TATAAGGCCACCAAGATGGGACGCTACATAAAGGATAGTGTGTCGCGGACAGGA 1380
Db 281975 TATAAGGCCACCAAGATGGGACGCTACATAAAGGATAGTGTGTCGCGGACAGGA 281916
QY 1381 AAAACAAATCTACTTGTAAACCCGACCTGAAGCCGGAAGAGCGTCACTATGAGGT 1440
Db 281915 AAAACAAATCTACTTGTAAACCCGACCTGAAGCCGGAAGAGCGTCACTATGAGGT 281856
QY 1441 GGGGTGTATACATTAACCCCGCGTCTGAATGCCAATGTACAGGTTTATGACTGAC 1500
Db 281855 GGGGTGTATACATTAACCCCGCGTCTGAATGCCAATGTACAGGTTTATGACTGAC 281796
QY 1501 TTCTCCAAACAAGATGTCTTATTCATAAATGATAACCAATAGCTATGTAAACAGC 1560
Db 281795 TTCTCCAAACAAGATGTCTTATTCATAAATGATAACCAATAGCTATGTAAACAGC 281736
QY 1561 GGAAGGCCCGGTTGACCGGTGGAATTTGCCGGCACAATGCGCGTGTGGTCAGAGAT 1620
Db 281735 GGAAGGCCCGGTTGACCGGTGGAATTTGCCGGCACAATGCGCGTGTGGTCAGAGAT 281676
QY 1621 GTACCGGTGCTACTGATTAACCTGACCGGAGTGAACAACGTCATGTGTATACAAA 1680
Db 281675 GTACCGGTGCTACTGATTAACCTGACCGGAGTGAACAACGTCATGTGTATACAAA 281616
QY 1681 GGTGCGCGCTGAGTTATACCCCTGAAACATCGTGAATGCGAAATGAACTGGCAGATC 1740

Db 281615 GGTGCGCGCTGAGTTATATCCCTGAACACATGTTGAATCGGAAACTGAACTGGCAGATC 281556
QY 1741 ACCGAAGAGTGGGCATCATCGCTGGGTGCGGTATTCGCGGGGAAACACACCGTTTCACC 1800
Db 281555 ACCGAAGAGTGGGCATCATCGCTGGGTGCGGTATTCGCGGGGAAACACACCGTTTCACC 281496
QY 1801 CAGAATTAATTCCTGCTCACTGAGCGCTGTACAGAAGAAAGTGTATGATGAGAAAGGAGATAC 1860
Db 281495 CAGAATTAATTCCTGCTCACTGAGCGCTGTACAGAAGAAAGTGTATGATGAGAAAGGAGATAC 281436
QY 1861 CTGAAACCTGGACGGTGGTGGATGACAGTCTGCTGTTGGAAGATGACGGATGCCCTGACG 1920
Db 281435 CTGAAACCTGGACGGTGGTGGATGACAGTCTGCTGTTGGAAGATGACGGATGCCCTGACG 281376
QY 1921 CTGAATGCTCGGTGAATAACCTGCTCAACAAGAAATTTACAGTACGCTGACGCTGTACAGT 1980
Db 281375 CTGAATGCTCGGTGAATAACCTGCTCAACAAGAAATTTACAGTACGCTGACGCTGTACAGT 281316
QY 1981 GCGGTAAGATGACGCTGTATGCGGTGATTTCTCCAGACGGGATCATCAACACAGGA 2040
Db 281315 GCGGTAAGATGACGCTGTATGCGGTGATTTCTCCAGACGGGATCATCAACACAGGA 281256
QY 2041 TATGTGATACCTGAGCGAAATTTACTGATGTCGCTGAACCTATCAGTTCTGA 2091
Db 281255 TATGTGATACCTGAGCGAAATTTACTGATGTCGCTGAACCTATCAGTTCTGA 281205

RESULT 9
AX702523/c 76804 bp DNA linear PAT 03-APR-2003
LOCUS AX702523 Sequence 109 from Patent WO02059320.
DEFINITION AX702523
ACCESSION AX702523
VERSION AX702523.1 GI:29537720
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE 1
AUTHORS Blattner, F.R., Welch, R.A. and Burland, V.D.
TITLE DNA sequences of Escherichia coli CPT073
PATENT: WO 02059320-A 109 01-AUG-2002;
JOURNAL WISCONSIN ALUMNI RESEARCH FOUNDATION (US)

FEATURES
 Location/Qualifiers
 1..76804
 /organism="Escherichia coli"
 /mol_type="unassigned DNA"
 /db_xref="taxon:562"
misc_feature 39751
 /note="Unsure"
misc_feature 49372
 /note="Unsure"
misc_feature 66539
 /note="Unsure"
misc_feature 73693
 /note="Unsure"
misc_feature 73705
 /note="Unsure"
misc_feature 73737
 /note="Unsure"
misc_feature 73759
 /note="Unsure"
misc_feature 73761
 /note="Unsure"
misc_feature 73786
 /note="Unsure"
misc_feature 73774
 /note="Unsure"
misc_feature 73781
 /note="Unsure"
misc_feature 73798
 /note="Unsure"
misc_feature 73802

ORIGIN									
/note="Unsure"									
Query Match 99.5%; Score 2079.8; DB 6; Length 76804;									
Best Local Similarity 99.7%; Pred. No. 0;									
Matches 2084; Conservative 0; Mismatches 7; Indels 0; Gaps 0;									
QY	1	ATGGGAATAACCACTCTGGCTTCGGTAGTCCGTATCATTCCTCTCGGATTTTCAGCCAGC	60						
DB	44573	ATGGGAATAACCACTCTGGCTTCGGTAGTCCGTATCATTCCTCTCGGATTTTCAGCCAGC	44514						
QY	61	ATAGCTGTCGAGAGGATGTGATGATTTCTCGGCATCCGGCTATGAGAAAAGCTGACT	120						
DB	44513	ATAGCTGTCGAGAGGATGTGATGATTTCTCGGCATCCGGCTATGAGAAAAGCTGACT	44454						
QY	121	AACGACGCCAGTGTTCGTGATTTAGCCAGGAGAAATGACATCCAGCCAGTACCAC	180						
DB	44453	AACGACGCCAGTGTTCGTGATTTAGCCAGGAGAAATGACATCCAGCCAGTACCAC	44394						
QY	181	GATCTGGCGGAGGCTCTGAGATCAGTAGAGGCTGTGATGCTGAAAGTGGTACGGGTAAA	240						
DB	44393	GATCTGGCGGAGGCTCTGAGATCAGTAGAGGCTGTGATGCTGAAAGTGGTACGGGTAAA	44334						
QY	241	ACCGGAGGCTGGAAATCAGATCCGAGGAATGCCAGCCAGTATACGCTGATCTGATT	300						
DB	44333	ACCGGAGGCTGGAAATCAGATCCGAGGAATGCCAGCCAGTATACGCTGATCTGATT	44274						
QY	301	GATGGTGTTCGTACGGCGGAAGCAGTGACGTGACTCCCAAGCGTTCCTGCGCATGAAT	360						
DB	44273	GATGGTGTTCGTACGGCGGAAGCAGTGACGTGACTCCCAAGCGTTCCTGCGCATGAAT	44214						
QY	361	ACCGGTCATGCCCCCTCTGGCGCCATTCAGCGTATTCAGGTTATACAGGGGCGCATG	420						
DB	44213	ACCGGTCATGCCCCCTCTGGCGCCATTCAGCGTATTCAGGTTATACAGGGGCGCATG	44154						
QY	421	TCCACATGTATGGCTCTGATGGATGGCGGTGTGGTGAATATATACCAAGAAAGAT	480						
DB	44153	TCCACATGTATGGCTCTGATGGATGGCGGTGTGGTGAATATATACCAAGAAAGAT	44094						
QY	481	GCAGACAAATGGCTCTCTCCGTCATATCGAGGCTGAATCTGAGGAAAGCAAAATGG	540						
DB	44093	GCAGACAAATGGCTCTCTCCGTCATATCGAGGCTGAATCTGAGGAAAGCAAAATGG	44034						
QY	541	GCTAACAGACCCAGTTTATTTCTGGAGCAGTGGTCCCTTGTGATGATCTGTACAG	600						
DB	44033	GCTAACAGACCCAGTTTATTTCTGGAGCAGTGGTCCCTTGTGATGATCTGTACAG	43974						
QY	601	CTGCAGGTACCGGTAGCACACACAGCGTCAGGGTTCATCGGTACATCACTGAGCGAT	660						
DB	43973	CTGCAGGTACCGGTAGCACACACAGCGTCAGGGTTCATCGGTACATCACTGAGCGAT	43914						
QY	661	ACAGCAGGACCGGTATTCCTTATCCACCGAGTCAAGAAATATATCTTGGTGCACGT	720						
DB	43913	ACAGCAGGACCGGTATTCCTTATCCACCGAGTCAAGAAATATATCTTGGTGCACGT	43854						
QY	721	CTTGACTGGAAGCGTCGGAGCAGGATGTCTCTGTTTGTATGATGATACCAACCGGAG	780						
DB	43853	CTTGACTGGAAGCGTCGGAGCAGGATGTCTCTGTTTGTATGATGATACCAACCGGAG	43794						
QY	781	CGTTATGATAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCCCTG	840						
DB	43793	CGTTATGATAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCCCTG	43734						
QY	841	CGCTATGAGCAACAAATTTTCAGTGGCTATGATCATACTTTACCTTCGGAAATCG	900						
DB	43733	CGCTATGAGCAACAAATTTTCAGTGGCTATGATCATACTTTACCTTCGGAAATCG	43674						
QY	901	AAATCGTATCTGAACCTGGAAACGAGACAGAAAAATAAAGGTCTGTACCGAGTGTGTA	960						
DB	43673	AAATCGTATCTGAACCTGGAAACGAGACAGAAAAATAAAGGTCTGTACCGAGTGTGTA	43614						
QY	961	CTGAAGCGCACAAATGGGGGCTTGGCGTACCGCGGAGCTTAAGAAATCGAACCTT	1020						
DB		CTGAAGCGCACAAATGGGGGCTTGGCGTACCGCGGAGCTTAAGAAATCGAACCTT							

Db	43613	CTGAAGCGCGACAAATGGGGGCTTGGCGTACAGCGCGGGAGCTTAAAGGAATCGAACCTT	43554
Qy	1021	ATCCTGAATTCATTACTGCTTACCCCTCTGGGAGAAATCTCATCTGGTTACCGTGGGGGC	1080
Db	43553	ATCCTGAATTCATTACTGCTTACCCCTCTGGGAGAAATCTCATCTGGTTACCGTGGGGGC	43494
Qy	1081	GAGTTTCAGAGCTCGTCCATGAAGAGCGAGTGTCTTGGCCAGCACACAGGTGAACATTTC	1140
Db	43493	GAGTTTCAGAGCTCGTCCATGAAGAGCGAGTGTCTTGGCCAGCACACAGGTGAACATTTC	43434
Qy	1141	CGGCAGAAAAGCTGGTCGGTATTTTGTCTGAGGATGAGTGGCATCTCACGGATGCACCTTCG	1200
Db	43433	CGGCAGAAAAGCTGGTCGGTATTTTGTCTGAGGATGAGTGGCATCTCACGGATGCACCTTCG	43374
Qy	1201	CTGACTGGGGCAGCGCGCTATGAAACATCATGAGCAATTCGGGGGACACTTCAGTCCGGCT	1260
Db	43373	CTGACTGGGGCAGCGCGCTATGAAACATCATGAGCAATTCGGGGGACACTTCAGTCCGGCT	43314
Qy	1261	GCATATCTGCTCTGGGATGTGGCAGATGCTCGACGCTGAAAGGCGGTGTGACACCGGA	1320
Db	43313	GCATATCTGCTCTGGGATGTGGCAGATGCTCGACGCTGAAAGGCGGTGTGACACCGGA	43254
Qy	1321	TATAAGGACCCAGAAATGGGCGAGCTACATAAAGGGAATTAGTGTGTGTCTCGGGCAGGGA	1380
Db	43253	TATAAGGACCCAGAAATGGGCGAGCTACATAAAGGGAATTAGTGTGTGTATCGGGCAGGGA	43194
Qy	1381	AAAAAATCTACTTGGTAAACCCCGACCTGAAAGCGGAGAGAGCGTCAGTTATGAGGCT	1440
Db	43193	AAAAAATCTACTTGGTAAACCCCGACCTGAAAGCGGAGAGAGCGTCAGTTATGAGGCT	43134
Qy	1441	GGGGTGTATTACGATAACCCCGCGGCTGTGAATGCCAATGTCAACAGGTTTTATGACTGAC	1500
Db	43133	GGGGTGTATTACGATAACCCCGCGGCTGTGAATGCCAATGTCAACAGGTTTTATGACTGAC	43074
Qy	1501	TTCTCCAAAGAGTTGTCTCTTATTTCCATAAATGATAACCAATAGCTATGTAACAGC	1560
Db	43073	TTCTCCAAAGAGTTGTCTCTTATTTCCATAAATGATAACCAATAGCTATGTAACAGC	43014
Qy	1561	GGAAAGCGCGGTTCACCGGTGGAATTTGCCGCGCACATTTGCCGCTGTGGTACAGGAT	1620
Db	43013	GGAAAGCGCGGTTCACCGGTGGAATTTGCCGCGCACATTTGCCGCTGTGGTACAGGAT	42954
Qy	1621	GTCAAGCTGTCACTGAATTAACCTGGACCGGAGTGAACAAAGTGTGATTAACAA	1680
Db	42953	GTCAAGCTGTCACTGAATTAACCTGGACCGGAGTGAACAAAGTGTGATTAACAA	42894
Qy	1681	GGTCCGCGCTGAGTTATACCCCTGAAACACATGTTGAATGCGAAACTGAACTGGCAGATC	1740
Db	42893	GGTCCGCGCTGAGTTATACCCCTGAAACACATGTTGAATGCGAAACTGAACTGGCAGATC	42834
Qy	1741	ACCGAAGAGGTGGCATCATGGTGGGTGCCGTTATTCGCGGAAACACACACGTTTCACC	1800
Db	42833	ACCGAAGAGGTGGCATCATGGTGGGTGCCGTTATTCGCGGAAACACACACGTTTCACC	42774
Qy	1801	CAGAAATATTTCCTCACTGAGCGCTGTACAGAGAAAGTGTATGATGAGAAAGGAGATAC	1860
Db	42773	CAGAAATATTTCCTCACTGAGCGCTGTACAGAGAAAGTGTATGATGAGAAAGGAGATAC	42714
Qy	1861	CTGAAAGCTGGACGCGTGGATGACAGTCTGTCTGTGAAAGATGACCGGATGCCCTGACG	1920
Db	42713	CTGAAAGCTGGACGCGTGGATGACAGTCTGTCTGTGAAAGATGACCGGATGCCCTGACG	42654
Qy	1921	CTGAATGCTGGGTGAATAACCTGCTCAACAAAGATTAACAGTACGTGAGCGCTGTACAGT	1980
Db	42653	CTGAATGCTGGGTGAATAACCTGCTCAACAAAGATTAACAGTACGTGAGCGCTGTACAGT	42594
Qy	1981	GCCGGTAAAGTACGCTGTATGCCGGTGATTAATCTTCCAGACGGGATCATCAACACAGGA	2040
Db	42593	GCCGGTAAAGTACGCTGTATGCCGGTGATTAATCTTCCAGACGGGATCATCAACACAGGA	42534
Qy	2041	TATGTGATACCTGAGCGAAATTAATCTGGATGTGCTGTGAACCTATCAGTTCTGA 2091	
Db	42533	TATGTGATACCTGAGCGAAATTAATCTGGATGTGCTGTGAACCTATCAGTTCTGA 42483	


```

RESULT 10
AE016766/c
LOCUS
DEFINITION
Escherichia coli CFT073 section 12 of 18 of the complete genome.
ACCESSION
AE016766 AE014075
VERSION
AE016766.1 GI:26109707
KEYWORDS
SOURCE
Escherichia coli CFT073
ORGANISM
Escherichia coli CFT073
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1 (bases 1 to 303121)
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
Extensive Mosaic Structure Revealed by the Complete Genome Sequence
of Uropathogenic Escherichia coli
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
12471157
2 (bases 1 to 303121)
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
Direct Submission
Submitted (20-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1. 303121
/organism="Escherichia coli CFT073"
/mol_type="genomic DNA"
/strain="CFT073"
/db_xref="taxon:199310"
177..1754
/gene="ygfU"
/locus_tag="c3465"
177..1754
/gene="ygfU"
/locus_tag="c3465"
1738..1881
/locus_tag="c3466"
1738..1881
/locus_tag="c3466"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AA081914.1"
/db_xref="GI:26109709"
/translation="MSMENNLLMRAIFSLLLFTLSIPVISDCVAMAIERSRPFYKM
IIF"
2004..2552
/locus_tag="c3467"
2004..2552
/locus_tag="c3467"
/EC_number="5.3.3.2"
/function="putative"
/notes="Escherichia coli K-12 ortholog: b2889; Escherichia
coli O157:H7 ortholog: z4227"
/codon_start=1
/transl_table=11
/product="isopentenyl-diphosphate delta-isomerase"
/protein_id="AA081915.1"
/db_xref="GI:26109710"
/translation="MQTEHVLINAGQVPTGLEKYAAHTADTLHLAFSSWLFNARG
OLIVTRALLSKAWPGWTFNSVCGHPOLGESNEAVIRCRVELGVEITPPEGIYPDF
RYATPDNGIENVGCPFAARTTSALQINDDEWMDYQNCDLADVLRGIDATPAWSP
WMMQATNREARKLSAFTQLX"
complement(2550..3260)
/insertion_seq="IS1541A-like"
complement(2657..3166)
/locus_tag="c3468"
complement(2657..3166)
/locus_tag="c3468"
/notes="Residues 18 to 169 of 169 are 94.07 pct identical
to residues 1 to 152 of 152 from Swissprot.40 :
>sp|Q57334|T200_SALTY_Transposase for insertion sequence
element IS200"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AA081916.1"
/db_xref="GI:26109711"
/translation="MRSGKCKQTRNKGQVPMGNEKSLAHTRWNCYHIVFAPYRRQ
VFYREKRAIGCILRLCEWKSVRILEACCADHIMLVEIIPKMSVSGFMVYKGS
SLMPYEQFGDLKPKYRNREFWCRGVYVDVVGKNTAKIQYIKHQBEDKMGQLSIPY
PGSPFTGRX"
complement(3306..4823)
/genes="lysS"
/locus_tag="c3469"
complement(3306..4823)
/genes="lysS"
/locus_tag="c3469"
/EC_number="6.1.1.6"
/function="enzyme; Aminoacyl tRNA synthetases, trna
modification"
/notes="Escherichia coli K-12 ortholog: b2890; Escherichia
coli O157:H7 ortholog: z4228"
/codon_start=1
/transl_table=11
/product="Lysyl-tRNA synthetase"
/protein_id="AA081917.1"
/db_xref="GI:26109712"
/translation="MSEQAQGADAVVDLNNELKTRREKANLREQIAPFDFRDH
TSQDLHAEPDGKNESELEALNIEVAVGRMTRRMKASFVTLQVGRGIQLYVARD
DLPEGVNPOFKWDLGDIILGAKGLFKTGELSHTCELRLTLKALRPLPKPHGL
QDOEARVORYLDLISNDSRNTFKVRSOILSGIROPMVNRGMEVETPMVOVPGGA
AAPEPTTHNALDLMYLRAPLRYLKLAVGSEFVEINFRNFRNEGISVRHNPEFT
MMELMAYADYKDLIELTSELFTLAQDLIGKTEVYGVDTLDGKPFPEKLTNREAIK
KXPEPTMADLDFNDSAKAIESIGIHVKSGLGRIVTEIFEVAAHLIIQITFTE
YPAEVPFLARNNDNPEITDRFEFFGGSEINGFSELDADQAOQLDQVAAKQAG
DDAMFYDEDYVTALREHGLPPTAGLGIGIDRMVMLFTNSHTTRDVLFPAMRPVK"
3423..4838
/locus_tag="c3470"
3423..4838
/locus_tag="c3470"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AA081918.1"
/db_xref="GI:26109713"
/translation="MFQCGDIIFVVEHRRFWVTCIFRGNLIOEALRLIFRVIQAKAVT
DFTPEELKAVCDFRNVVVTQRQNFRRFNSRGLRQDMCFEFDLGNDAQIIP
AFNLVADARFSNRRFVEVQIRHVSFAIFLDRFTHGQGFERTVEVQYVAGHFRLL
TENILQOSTEQRFQGLDKIFVICVSHIEFHGELWMTYGTNTFTVTVDFYTLLEAT
DHQTLFVQFRRDTQVHVEIQRMVMSDKRTRSGTARDHLHRRRPNFKAAVHLEANTR

```



```
EDLRAHFKGVAGSIVRDEIQITLTIARFLILOAVEFIORTOCFQOQTOLGTVDRPFA
GGFGEFTTRADIEABVPFFELLVYNAFQIVTGNVQLNATANVLQNERPETHNTG
HAAKNGDFDQRFQFVLFTFFVFCVQLVRGMIAIEIVREGNPLLTQVRLTLTRPQF
IKVIDYRSALCVLFRHVGSS"
complement(4833..5931)
/gene="prfB"
/gene="prfB"
/locus_tag="c5622"
complement(join(4833..5855,5857..5931))
/gene="prfB"
/locus_tag="c5622"
/function="factor; Proteins - translation and
modification"
/notes="RF-2; Escherichia coli K-12 ortholog: b2891;
Escherichia coli O157:H7 ortholog: z4229; ribosomal
slippage: an in-frame UGA termination codon is located
within the prfB sequence, and a naturally occurring +1
frameshift is required for synthesis of RF-2"
/codon_start=1
/transl_table=11
/product="Peptide chain release factor 2"
/protein_id="AAN8191.1"
/db_xref="GI:26109714"
/translation="MFEINPNVRQDLTERSDFVLRGVLDYDAKERLEEYNABLEQP
DVNPERAQAQKERSSEUAVDVLDDQMKQLEDVSGLELAVADDETFNEKVAE
LDALEKLAQLFRMFSGEYSDCYLDIQAGSGGTBAQDWASMLERMYLRWASRG
FTEIIESEGEVAGIKSVTIKISGDYAGWLRTEGVRLVRKSPFDSGRRHTFS
SAFVPEVDDIDIEINPADLRIDVVRASGAGGQHVNRTESSAVRITHPTGIVTQCN
DRSQHNKQOAMKAKLYELEMOKNAEKQAMEDNKSDDIGWSQIRSYVLDSDRIK
DURTGVETRNQVLDGSDQIEASLKAGL"
complement(6022..7755)
/gene="recJ"
/locus_tag="c3472"
complement(6022..7755)
/gene="recJ"
/locus_tag="c3472"
/EC_number="3.1.-.-"
/function="enzyme; Degradation of DNA"
/notes="Escherichia coli K-12 ortholog: b2892; Escherichia
coli O157:H7 ortholog: z4230"
/codon_start=1
/transl_table=11
/product="Single-stranded-DNA-specific exonuclease recJ"
/protein_id="AAN8192.1"
/db_xref="GI:26109715"
/translation="MKQIQILRRREVDETDLPABELPRLRLYASRGVRSQAHLERS
VKGMLPQQLSGVEKAVELLYNAFREGRITLIVVDFDADGATSTALSVMRSLGCSN
IDYLVNRPEDGYSPEYVDAQHARGAQLIVTVDNGISSHAGVEHARSLSGIPVITD
HILPGDTLPAABAIINPNLDCNFPKSLAGVGAFYLMALRTFLRQGWPFDERGIA
IPNALLEDLVALGTADVVDLNNRILTWQMSRIIRAGKCRPGIKALLEVANDRPIQ
KLAASDLGALGPLNLAAGLDDMSGVALLLCDNIGEARVLANSLDLALNOTRKEIQ
GMQVEALTCEKLRSDRLPGLGLAMYPHEWQGVVGLASRIKEFRHPRVIAFAPAG
DGTGKSGRSIQGLNHRDALERLDLYLPGMLKFGHAWAAGLSLEBDFELFQRFQ
ELVTEWLDSPLSQGVSDGSLPAAEMTEVAQLLRDGPWQMFEPFLFGHFLLQ
QLVLVGRHLKVMVEPVGGGFLDDIAFNVDLTALMPDNGVREVLAYKLDINEPRGNS
LQIIIDNIWPI"
complement(7761..8471)
/gene="dsbC"
/locus_tag="c3473"
Query Match 99.5%; Score 2079.8; DB 1; Length 303121;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2084; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 ATCGGAATACCACTCGGTCCTCGTAGTCATTCCTCGTCTCGGATTTTCACGCCAGCAGC 60
DB 123542 ATCGGAATACCACTCGGTCCTCGTAGTCATTCCTCGTCTCGGATTTTCACGCCAGCAGC 123483
QY 61 ATAGCTGCTCAGAGGATGTGATGATGTCTCGGCATCCGGTATCGAGAAAAGCTGACT 120
DB 123482 ATAGCTGCTCAGAGGATGTGATGATGTCTCGGCATCCGGTATCGAGAAAAGCTGACT 123423
QY 121 AAGCGAGCCGCGAGTGTCTGTGATAGCAGGAGGAATTCAGTCCAGCCAGTACCAC 180
|||||
```

```
123422 AACGACGCCCCAGTGTCTCTGATTAGCCAGGAGGAATTGCAGTCCAGCCAGTACCAC 123363
QY 181 GATCTGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGATGTTCGAAAGTGTGATCGGTAAA 240
123362 GATCTGGCGGAGGCTCTGAGATCAGTAGAGGGTGTGATGTTCGAAAGTGTGATCGGTAAA 123303
QY 241 ACCGGAGGGCTGGAAATCAGCATCCGAGGAATCCAGCCAGGTTACACGCTGATCTGATT 300
123302 ACCGGTGGCTGGAAATCAGCATCCGAGGAATCCAGCCAGGTTACACGCTGATCTGATT 123243
QY 301 GATGGTGTCTGTCAGGCGGAAGCAGTGCCTGACTCCCAACGGTTTTCTGTCCTCATGAAT 360
123242 GATGGTGTCTGTCAGGCGGAAGCAGTGCCTGACTCCCAACGGTTTTCTGTCCTCATGAAT 123183
QY 361 ACCGGGTTCATGCCCTCTGTCGCGCCATTGAGCGGTATTGAGGTTATCAGGGGGCCGATG 420
123182 ACCGGGTTCATGCCCTCTGTCGCGCCATTGAGCGGTATTGAGGTTATCAGGGGGCCGATG 123123
QY 421 TCCACACTGTATGGCTCTGATGCGATGGCGGTGTGGTGAATATCATATTACCAGAAAGAT 480
123122 TCCACACTGTATGGCTCTGATGCGATGGCGGTGTGGTGAATATCATATTACCAGAAAGAT 123063
QY 481 GCAGACAAATGGCTCTCTCCGTCATGTCAGGGCTGAATCTGCAGGAAAGCAACAATGG 540
123062 GCAGACAAATGGCTCTCTCCGTCATGTCAGGGCTGAATCTGCAGGAAAGCAACAATGG 123003
QY 541 GGTAACAGCAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCTTGTGGATGATTTCTGTGACG 600
123002 GGTAACAGCAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCTTGTGGATGATTTCTGTGACG 122943
QY 601 CTGCAGGTACGCGGTAGCACACACACGCTCAGGTTTCATCGTTCATCATCATCATGACGCGAT 660
122942 CTGCAGGTACGCGGTAGCACACACACGCTCAGGTTTCATCGTTCATCATCATCATGACGCGAT 122883
QY 661 ACAGCAGCAGCGGTATTCTCTTATCCACGGAGTTCACAGAAATTAATCTTGTGTGACCGT 720
122882 ACAGCAGCAGCGGTATTCTCTTATCCACGGAGTTCACAGAAATTAATCTTGTGTGACCGT 122823
QY 721 CTTCAGTGGAGGCTCGGAGGATGTCTCTGTTTGTATGATGATACCCCGGCGAG 780
122822 CTTCAGTGGAGGCTCGGAGGATGTCTCTGTTTGTATGATGATACCCCGGCGAG 122763
QY 781 CGTTATCATAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCCCTG 840
122762 CGTTATCATAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCCCTG 122703
QY 841 CGCTATGAGCGAAACAAAATTTGAGTGGCTATGATCATCTTTTACCTTCGGAAACATGG 900
122702 CGCTATGAGCGAAACAAAATTTGAGTGGCTATGATCATCTTTTACCTTCGGAAACATGG 122643
QY 901 AAATCGTATCTGAACGTGAACGAGACAGAAAAATAAGGTCGTGAGCTTGTACCGAGTGA 960
122642 AAATCGTATCTGAACGTGAACGAGACAGAAAAATAAGGTCGTGAGCTTGTACCGAGTGA 122583
QY 961 CTGAAGCGCGACAAATGGGGGCTTTCGGGTGACGCGGGGAGCTTAAGGAATCGAACCTT 1020
122582 CTGAAGCGCGACAAATGGGGGCTTTCGGGTGACGCGGGGAGCTTAAGGAATCGAACCTT 122523
QY 1021 ATCCTGAATTCATTACTGCTTACCCCTCTGGGAGATCTCATCTGTTACGTTGGGGGGC 1080
122522 ATCCTGAATTCATTACTGCTTACCCCTCTGGGAGATCTCATCTGTTACGTTGGGGGGC 122463
QY 1081 GAGTTTCAGAGCTCGTCCATGAAAGACGGAGTTGTCTCTCCAGCAGCAGGTGAAACTTTC 1140
122462 GAGTTTCAGAGCTCGTCCATGAAAGACGGAGTTGTCTCTCCAGCAGCAGGTGAAACTTTC 122403
QY 1141 CGGCAGAAAAGCTGTCGTTATTGTCGAGGATGAGTGGCATCTCAGGATGACCTTGG 1200
122402 CGGCAGAAAAGCTGTCGTTATTGTCGAGGATGAGTGGCATCTCAGGATGACCTTGG 122343
QY 1201 CTGACTCGGGGACGCCCTATGAACATCATGAGCAATTCGGGGGACACTTTCAGTCCGCGT 1260
122342 CTGACTCGGGGACGCCCTATGAACATCATGAGCAATTCGGGGGACACTTTCAGTCCGCGT 122283
DB
```


QY 454 GTGGTGAATATCATTAACAGAAAGAAATCGACAAATGGCTCTCTTCGGTCAATGCAGG 513
 DB 13290 GTGGTGAATATCATTAACAGAAAGAAATCGACAAATGGCTCTCTTCGGTCAATGCAGG 13231
 QY 514 CTGAATCTGAGGAAAGCAAAATGGGGTAAACAGCAGCCAGTTTAAATTTCTGAGCAGT 573
 DB 13230 CTGAATCTGAGGAAAGCAAAATGGGGTAAACAGCAGCCAGTTTAAATTTCTGAGCAGT 13171
 QY 574 GGTCCCTTGTGGATGATCTGTGACCTCGAGGTACGGGTAGCACAACAGGTCAG 633
 DB 13170 GGTCCCTTGTGGATGATCTGTGACCTCGAGGTACGGGTAGCACAACAGGTCAG 13111
 QY 634 GGTTCATCGGTACATCACTGAGCGATACAGCAGCAGCGGTATTCCTTATCCACCGAG 693
 DB 13110 GGTTCATCGGTACATCACTGAGCGATACAGCAGCAGCGGTATTCCTTATCCACCGAG 13051
 QY 694 TCACAGAAATTAATCTTGGTGCACGTCTGTGACTGGAAGGGCTGCGAGCAGATGTGCTC 753
 DB 13050 TCACAGAAATTAATCTTGGTGCACGTCTGTGACTGGAAGGGCTGCGAGCAGATGTGCTC 12991
 QY 754 TGGTTTGATATGATACCAACCGGAGCGGTATGATTAACCGGGATGGGCAACTGGGGAGT 813
 DB 12990 TGGTTTGATATGATACCAACCGGAGCGGTATGATTAACCGGGATGGGCAACTGGGGAGT 12931
 QY 814 CTGACGGGGGATATGACCGGACCTCGCTATGACGGAACAAATTTGAGCTGGCTAT 873
 DB 12930 CTGACGGGGGATATGACCGGACCTCGCTATGACGGAACAAATTTGAGCTGGCTAT 12871
 QY 874 GATCATCTTTACCTTCGGAACATGGAATCGTATCTGAACTGGAACGAGCAGAAAT 933
 DB 12870 GATCATCTTTACCTTCGGAACATGGAATCGTATCTGAACTGGAACGAGCAGAAAT 12811
 QY 934 AAAGTCTGAGCTTGTACCGAGTGTACTGAAGCGCGCAAAATGGGGCTGCGGGTCAG 993
 DB 12810 AAAGTCTGAGCTTGTACCGAGTGTACTGAAGCGCGCAAAATGGGGCTGCGGGTCAG 12751
 QY 994 CCGGGAGCTTAAGAAATCGAAATCTATCTGAAATTCATCTACTCTTACCCCTCGGA 1053
 DB 12750 CCGGGAGCTTAAGAAATCGAAATCTATCTGAAATTCATCTACTCTTACCCCTCGGA 12691
 QY 1054 GAATCTCATCTGTTACGGTGGGGGAGTTTTCAGAGCTCTGCTCAAGAGCAGGTT 1113
 DB 12690 GAATCTCATCTGTTACGGTGGGGGAGTTTTCAGAGCTCTGCTCAAGAGCAGGTT 12631
 QY 1114 GTCTTTCGACGACAGGTGAATCTTCCGGCAGAAAGCTGCTGGTATTTGCTGAGAT 1173
 DB 12630 GTCTTTCGACGACAGGTGAATCTTCCGGCAGAAAGCTGCTGGTATTTGCTGAGAT 12571
 QY 1174 GAGTGCATCTCAGGATCACTTGGCTGACTGCGGAGCGGCTATGAAATCATGAG 1233
 DB 12570 GAGTGCATCTCAGGATCACTTGGCTGACTGCGGAGCGGCTATGAAATCATGAG 12511
 QY 1234 CAATTCGGGGGACATTCAGTCCGGTGCATATCTGGTCTGGGATGTGGCAGATCCCTGG 1293
 DB 12510 CAATTCGGGGGACATTCAGTCCGGTGCATATCTGGTCTGGGATGTGGCAGATCCCTGG 12451
 QY 1294 AGCTGAAGAGCGGTGTGACACGGGATTAAGGACCCAGAAATGGGGCAGTACATAA 1353
 DB 12450 AGCTGAAGAGCGGTGTGACACGGGATTAAGGACCCAGAAATGGGGCAGTACATAA 12391
 QY 1354 GGGATTAGTGTGTGTCGGGAGGGAACAAATCTACTTGTGTAACCCGACCTGAAG 1413
 DB 12390 GGGATTAGTGTGTATCCGGGAGGGAACAAATCTACTTGTGTAACCCGACCTGAAG 12331
 QY 1414 CCGGAAGAGAGCGTCAAGTTATAGGCTGGGGTGTATTACGATAACCCCGCGGTCTGAAT 1473
 DB 12330 CCGGAAGAGAGCGTCAAGTTATAGGCTGGGGTGTATTACGATAACCCCGCGGTCTGAAT 12271
 QY 1474 GCAATGTCAAGTTTATGACTGACTCTCCACAGATGTCTTATTCCATAAT 1533
 DB 12270 GCAATGTCAAGTTTATGACTGACTCTCCACAGATGTCTTATTCCATAAT 12211

QY 1534 GATAACACCAATAGCTATGTAAACAGCGGAAGCCCGGTTCACGGTGTGGAATTTGCC 1593
 DB 12210 GATAACACCAATAGCTATGTAAACAGCGGAAGCCCGGTTCACGGTGTGGAATTTGCC 12151
 QY 1594 GGACATATGCGCTGTGGTTCAGAGGATGTACGCTGTCACTGAATTTACACCTGGACCGA 1653
 DB 12150 GGACATATGCGCTGTGGTTCAGAGGATGTACGCTGTCACTGAATTTACACCTGGACCGA 12091
 QY 1654 AGTGAACCAACGTGATGGTGATAACAAAGGTGCGCCGCTGAGTTTATACCCCTGAACACATG 1713
 DB 12090 AGTGAACCAACGTGATGGTGATAACAAAGGTGCGCCGCTGAGTTTATACCCCTGAACACATG 12031
 QY 1714 GTGAATCGGAAATGAACTGGCGAGATCAACGAAAGGTGCGCATCATGGTGGGTGCGCGT 1773
 DB 12030 GTGAATCGGAAATGAACTGGCGAGATCAACGAAAGGTGCGCATCATGGTGGGTGCGCGT 11971
 QY 1774 TATCGGGGAAACACACCGTTTACCCAGAAATTTATTCGTCACTGAGCGCTGTACAGAAG 1833
 DB 11970 TATCGGGGAAACACACCGTTTACCCAGAAATTTATTCGTCACTGAGCGCTGTACAGAAG 11911
 QY 1834 AAAGTGTATGATGAGAAAGAGAAATACCTGAAAGCTGGAACGCTGTGATGATGAGTCTG 1893
 DB 11910 AAAGTGTATGATGAGAAAGAGAAATACCTGAAAGCTGGAACGCTGTGATGATGAGTCTG 11851
 QY 1894 TCGTGGAGATGAGGATGCGCTGACGCTGAATGCTGCGGTGAATTAACCTGCTCAACAAG 1953
 DB 11850 TCGTGGAGATGAGGATGCGCTGACGCTGAATGCTGCGGTGAATTAACCTGCTCAACAAG 11791
 QY 1954 GATTACAGTCACTGAGCCCTGTACAGTCCGGTGAAGAGTACGCTGTATGCGGTGATTAC 2013
 DB 11790 GATTACAGTCACTGAGCCCTGTACAGTCCGGTGAAGAGTACGCTGTATGCGGTGATTAC 11731
 QY 2014 TTCAGACGGGATCATCAACAGAGATGTATGATACCTGAGCGAAATTTACTGATGTCG 2073
 DB 11730 TTCAGACGGGATCATCAACAGAGATGTATGATACCTGAGCGAAATTTACTGATGTCG 11671
 QY 2074 CTGAATCATCAGTTCTGA 2091
 DB 11670 CTGAATCATCAGTTCTGA 11653

RESULT 12
 ECO278144 37710 bp DNA linear BCT 30-NOV-2001
 LOCUS
 DEFINITION Escherichia coli partial genomic DNA for pathogenicity island,
 shiga toxin producing strain 4797/97.
 ACCESSION
 VERSION AJ278144.1 GI:14252925
 KEYWORDS
 int gene; integrase; ORF10; ORF11; ORF12; ORF13; orf15; ORF16;
 ORF17; ORF18; ORF19; ORF4; ORF5; ORF6; ORF7; ORF8; ORF9; putative
 sodium galactoside symporter; yicJ gene.
 SOURCE
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE
 1 Schmidt, H., Zhang, W.L., Hemmrich, U., Jelacic, S., Brunder, W.,
 Tarr, P.I., Dobrindt, U., Hacker, J. and Karch, H.
 Identification and characterization of a novel genomic island
 integrated at selC in locus of enterocyte effacement-negative,
 Shiga toxin-producing Escherichia coli
 Infect. Immun. 69 (11), 6863-6873 (2001)
 JOURNAL
 MEDLINE 21481963
 PUBMED 11598060
 REFERENCE
 2 (bases 1 to 37710)
 Zhang, W.L.
 Direct Submission
 Submitted (22-MAY-2000) Zhang W.L., Institut fuer Hygiene und
 Mikrobiologie, Universitaet Wuerzburg, Josef-Schneider-Str. 2,
 D-97080, GERMANY
 FEATURES
 source
 1..37710
 /organism="Escherichia coli"

CDS

17821..19908
/function="adhesion"
/note="ORF9"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC39288.1"
/db_xref="GI:14252933"
/db_xref="GOA:O93SE1"
/db_xref="SPTREMBL:O93SE1"
/translation="NRITLASVIPCIFPSASSMAEDVMIVSASGVKKLTNAAS
VSVISBELQSQHDLAEARAVEGVDSGTGDLGLGISIRGMPASYLLILIDGV
RQGGSDVTPNGSAMNTGFMPPPLAAIERIEVIFPMSTLYSGDAMGVNIIITKKA
DKWLSNVAGNLQSNKWSQFNFWSGLPVDSDLSLQIRGSTQQRQGSVTSLS
DTSATRIPIYPTESQNYNLGARIDWKATQDQVDMFTTRQYDNDQQLGSLTGYSID
HLTVRNIKSAGYDHTTFTGKSYLNWETENKRELVSLSKDKWGLAGPREL
KSNLLINSLLPLTPEGSHLLTVGGEYSQSSMKQGVLAISGTENPROKSWFAEDWE
HUTDVALTAGRYRHHQPFQGHFSPRAYLVVDVTDANTLKGVTTGYKATMQLK
GISVGGQGRNLLNPELPEESVYEGVYDNLGNLNTGTFIDFNKLVSYL
IDNTNSYTNSSKARLHGVFAGTLPMSDEVTLSLNTWTQSEQRDNDKGAFLSYT
PEHWNALNMQITBEVASWLGARYRGTPTFRNYSLSVQKRVYDEKEYLKAWT
VVDAGLSWKVTDALTLNAAVNNLLNKDYSVLSYADSKSTLYAGDYFQGSSTTGYVI
PERNIWMSLNTQF"
complement(20254..22254)

CDS

Query Match 89.1%; Score 1863.8; DB 1; Length 37710;
Best Local Similarity 93.5%; Pred. No. 0;
Matches 1956; Conservative 0; Mismatches 132; Indels 3; Gaps 1;
QY 1 ATGCGAATAACCACTCGCTCCGTTCCGTTAGTCATCCCTGTCGGAATTTTCAGCCAGCAGC 60
DB 17821 ATGCGAATAACCACTCGCTCCGTTCCGTTAGTCATCCCTGTCGGAATTTTCAGCCAGCAGC 17880
QY 61 ATAGCTGTCGACAGAGATGTGATGTTGTCGCGCATCCGGCTATGAGAAAAGCTGACT 120
DB 17881 AT---GGCAGCAGAGATGTGATGTTGTCGCGCATCCGGCTATGAGAAAAGCTGACT 17937
QY 121 AACGACGCCAGCTGTTTCTGTGATTAGCCAGGAGGANTTCAGTCCAGCCAGTACCAC 180
DB 17938 AACGACGCCAGCTGTTTCCGTAATCAGTCAGGAGGAAATGCAGTCAGCCAGTACCAC 17997
QY 181 GATCTGGCGAGGCTCTGAGATCAGTAGAGGTGTGATGTTGAAAGTGGTACGGGTAAA 240
DB 17998 GATCTGGCGAGGCACTGAGAGCTGTGGAGGTGTGATGTTGAAAGTGGTACGGGTAAA 18057
QY 241 ACCGAGGAGGTGAAATACGATCCAGGAATCCAGCCAGTTCACCGTGTATCTGATT 300
DB 18058 ACCGAGGAGGTGAAATACGATCCGTTGAATCCCGCAGTTCACCGTGTATCTGATT 18117
QY 301 GATGTTGTCGTCAGCGCGGAGCAGTGCAGTCCCAACGGTTTTTCTGCCATGAAT 360
DB 18118 GATGTTGTCGTCAGCGCGGAGCGGAGCGTGCAGTCCCAACGGTTTTTCTGCCATGAAT 18177
QY 361 ACCGGGTTTCATGCCCTCTGGCGCCATTGAGCGTATTGAGTTATCAGGGGGCGGATG 420
DB 18178 ACAGGGTTTCATGCCCTCTGGCGCCGATTGAGCGTATTGAGTTATCAGGGGGCGGATG 18237
QY 421 TCACACTGTATGGCTGTGATGCGATGGGCGGTGTGTTGAATATCATTCACAGAGGAT 480
DB 18238 TCACACTGTATGGCTGTGATGCGATGGGCGGTGTGTTGAATATCATTCACCAAGGAT 18297
QY 481 GCAGACAATGGCTCTCTCCGTCATGCGAGGCTGAATCTGCAGGAAACCAAAATGG 540
DB 18298 GCAGACAATGGCTCTCTCCGTCACCGCGGCTGAATCTGCAGGAGATGAACATGG 18357
QY 541 GGTAAACAGCAGCAGTTTAAATTTCTGGAGCAGTGGTCCCTGTGTGATGATTTCTGTACG 600
DB 18358 GGCAACAGCAGCAGTTTAAATTTCTGGAGCAGTGGTCCCTGTGTGATGATTTCTGTACG 18417
QY 601 CTCAGGTACCGGTAGCAGACACAGCGTCAGGTTTCATCGTTCATCATCTGACGCGAT 660
DB 18418 CTCAGATCCCGGCGAGCACACAGCAGCGTCAGGTTTCATCGTTCATCATCTGACGCGAC 18477

QY 661 ACAGCAGGCGACGGTATTCTTATCCACGAGTCAAGAAATTAATCTTGTGTGACGT 720
DB 18478 ACATCAGCCACGGTATTCTTATCCACGAGTCAAGAAATTAATCTTGTGTGCGGT 18537
QY 721 CTTGACTGGAAGCGTCGAGCAGAGTGTCTCTGTTTGTATGATGATACACCCGCGCAG 780
DB 18538 ATTGACTGGAAGCGTCGAGCAGAGTGTCTCTGTTTGTATGATGATACACCCGCGCAG 18597
QY 781 CGTTATGATAACCGGGATGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCGCTG 840
DB 18598 CGTTATGATAACCGGGATGGTCAACTGGGTAGCTGACAGGGGGATATGACCACTACTG 18657
QY 841 CGCTATGAGGAAACAAAATTTGAGTGGCTATGATCATATCTTCACTTCGGAACATGG 900
DB 18658 CGCTATGAGGAAACAAAATTTGAGTGGCTATGATCATATCTTCACTTCGGTACTGCTG 18717
QY 901 AATTCGTATCTGAATCGAACGAGACAGAAAATAAAGTCTGAGCTTTGACGAGTGTGTA 960
DB 18718 AATTCGTATCTGAATCGAACGAGACAGAAAATAAAGTCTGAGCTTTGACGAGTGTGTA 18777
QY 961 CTGAAGCGCAGAAATGGGGCTTCCGCTGACCGCGGGAGCTTAAGGAATGAAACCTT 1020
DB 18778 CTGAAGCGCAGAAATGGGGCTTCCGCTGAAACCGCGGGAGCTTAAGGAATGAAACCTT 18837
QY 1021 ATCCTGAATTCATTACTGCTTACCCCTCTCGGAGAAATCTCATCTGTTTACGGTGGGGGC 1080
DB 18838 ATCCTGAATTCATTACTGCTTACCCCTCTCGGAGAAATCTCATCTGTTTACGGTGGGGGC 18897
QY 1081 GAGTTTCAGAGCTCGTCCATGAAAGACGAGTGTGTCCTTCCAGCAGCAGGTGAACTTTC 1140
DB 18898 GAGTTTCAGAGCTCGTCCATGAAAGACGAGTGTGTCCTTCCAGCAGCAGGTGAACTTTC 18957
QY 1141 CGGACAGAAAAGCTGGTGGTATTTGCTGAGGATGAGTGGCATCTCAGGGATGCACTTTCG 1200
DB 18958 CGACAGAAAAGCTGGTGGTATTTGCTGAGGATGAGTGGCATCTCAGGGATGCACTTTCG 19017
QY 1201 CTGACTCGGGCAGCCGCTATGAACATCATGAGCAATTCGGGGACACTTCAGTCCCGCT 1260
DB 19018 CTGACTCGGGCAGCTCGTATGAACATCATGATCAATTCGGGGACACTTCAGTCCCGCT 19077
QY 1261 GCATATCTGTTCTGGGATGTGGCAGATGCTGACGCTGAAAGCGGTGTGACCCAGGGA 1320
DB 19078 GCATATCTGTTCTGGGACGTGACCGATGCTGGAAGCGGTGTGACCCAGGGA 19137
QY 1321 TATAAGCACCACCAATGGGGCAGCTACATAAAGGATTAGTGTGTCTCGGGCAGGGA 1380
DB 19138 TATAAGCCCCCACCATTGGGACAGTTACATAAGGGATCAGCGTGTATCCGGCAGGGA 19197
QY 1381 AAAACAAATCTACTTGTGTAACCCCGACTGAAGCCGGAAGAGAGTGAAGTTATGAGCG 1440
DB 19198 CGTACTAATCTACTCGGTAAACCCCGACTGAAGCCGGAAGAGAGTGAAGTTATGAGCG 19257
QY 1441 GGGTGTATTACGATAACCCCGCGGTCTGAATGCCAATGTACAGAGTTTATGACTGAC 1500
DB 19258 GGGGTATTATTACGATAACCTCTCCGGTCTGAATGCCAATTAACAGGCTTTTATACAGAC 19317
QY 1501 TTCTCCAAACAGATTGTCTTATTTCATTAATGATAACCAATAGCTATGTAACAGC 1560
DB 19318 TTCTCCAAACAGATTGTCTTATTTCATTAATGATAACCAATAGCTATGTAACAGC 19377
QY 1561 GGAAGGCCCGGTGACAGGTGTGGAATTTCCCGGACATTGCCGCTGTGTGTGAGAGAT 1620
DB 19378 GGAAGGCCCGGTGACAGGTGTGGAATTTCCCGGACATTGCCGCTGTGTGTGAGAGAT 19437
QY 1621 GTCACGCTGTCACTGAATTAACCTGGACCCGAGTGAACAACTGATGGTGAATAACAAA 1680
DB 19438 GTCACCTGTCTGTGAAATTAACCTGGACCCGAGTGAACAACTGATGGTGAATAACAAA 19497
QY 1681 GGTGGCGCGCTGAGTTATACCCCTGAAACACATGTTGTAATGCGAAACTGAACCTGCGAGT 1740
DB 19498 GGTGACCGCTGAGTTATACCCCTGAAACACATGTTGTAATGCGAAACTGAACCTGCGAGT 19557
QY 1741 ACCGAGAGGTGGCATCATGCTGGTGGCTTATCGCGGAAAAACACCACTGTTTACC 1800


```

Db 19558 ACCGAAGTGGGATCATGGCTGGGTCGCCGTTATCGCGGAAACACACAGTTTCACC 19617
Qy 1801 CAGAAATTATTCCTCAGCTGAGCGCTGTACAGAAAGAGTGTATGATGAGAAAGAGATAC 1860
Db 19618 CGGAATTATTCCTGCTGAGCGCTGTACAGAAAGAGTGTATGATGAGAAAGAGATAT 19677
Qy 1861 CTGAAAGCTGGAGCGGTGGATGAGCTGTGCTGCTGGAGATGACGGATGCCCTGACG 1920
Db 19678 CTGAAAGCTGGAGCGGTGGATGAGCTGTGCTGCTGGAGATGACGGATGCCCTGACG 19737
Qy 1921 CTGAATGCTGGGTGAATAACCTGCTCAACAAGATTACAGTACAGTACAGCTGTACAGT 1980
Db 19738 CTGAATGCTGGGTGAATAACCTGCTCAACAAGATTACAGTACAGTACAGCTGTACAGT 19797
Qy 1981 GCCGGTAAGAGTACGCTGTATGCGGTGATTACTTCAGACGGGATCATCAACACGGA 2040
Db 19798 GCCGATAAGAGTACGCTTTATGCGGTGATTACTTCAGACGGGATCATCAACACGGA 19857
Qy 2041 TATGTGATACCTGAGCGAATTAATCTGATGCTGCTGCTGAATCATGCTCTGA 2091
Db 19858 TATGTGATACCTGAGCGAATTAATCTGATGCTGCTGCTGAATCATGCTCTGA 19908

RESULT 13
AF447814 71684 bp DNA linear BCT 16-OCT-2002
LOCUS Escherichia coli strain CFT073 pathogenicity island II, partial
DEFINITION sequence.
ACCESSION AF447814
VERSION AF447814.1 GI:18265859
KEYWORDS Escherichia coli
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 71684)
AUTHORS Rasko, D.A., Phillips, J.A., Li, X. and Mobley, H.L.
TITLE Identification of DNA sequences from a second pathogenicity island
of uropathogenic Escherichia coli CFT073: probes specific for
uropathogenic populations
J. Infect. Dis. 184 (6), 1041-1049 (2001)
MEDLINE 21458620
PUBMED 11574920
REFERENCE 2 (bases 1 to 71684)
AUTHORS Rasko, D.A., Phillips, J.A., Li, X. and Mobley, H.L.T.
TITLE Direct Submission
JOURNAL Submitted (14-NOV-2001) Dept of Microbiology and Immunology,
University of Maryland School of Medicine, 655 W. Baltimore Street,
Baltimore, MD 21201, USA
FEATURES
Location/Qualifiers
1..71684
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="CFT073"
/db_xref="taxon:562"
/notes="pathogenicity island II; isolated from the blood of
a Homo sapiens female; urinary tract pathogen"
<1..339
/notes="ORF1; similar to Escherichia coli L0014 encoded by
GenBank Accession Number AF071034"
/codon_start=1
/transl_table=11
/product="L0014-like protein"
/protein_id="AAL67333.1"
/db_xref="GI:18265860"
/translation="GSHSHOVPAAGSWPALPICVNPSTWENYNNCWMPSPSVTCL
SFVADGVATLLKYSGPLMYCASPNAWRKASLSGLRCVTYVPLPARNWQCSSISWTG
VSQKHSALTH"
334..1028
/notes="ORF2; similar to Escherichia coli L0015 encoded by
GenBank Accession Number AF071034"
/codon_start=1

```

```

/transl_table=11
/product="L0015-like protein"
/protein_id="AAL67334.1"
/db_xref="GI:18265861"
/translation="MPGRIITKMSQKYLIRIAFLESQKQKQOLSLVETKTFLLRS
ALAPAEKIDEDDEIEHLRAQIEKRLMFGTCEKLRREVEQAELAKQROSDSPR
YSGREDPQPRQLRQSEHRRPPFAHLPREIHRLESSESCPCGCGELDYLGVSAEQ
LELVSSALKVIRTERVKKACTKDCICIVEAPAPRRDRAWYGRPRITTCPRVYKILLRTP
ATVSSSE"
1102..1929
/notes="ORF3; similar to Escherichia coli L0015 encoded by
GenBank Accession Number AF071034"
/codon_start=1
/transl_table=11
/product="L0015-like protein"
/protein_id="AAL67335.1"
/db_xref="GI:18265862"
/translation="MTPLNDALYRYVMNTRKLHTDDTPVKVLAFLPGLKTKTGRITVY
RDDNAGSSPPAVPVPAYSPNROGKHPQHLRPPGFIQADATGYDRLPSARREGGA
LTVACCAHARRKTHDVYISKSNMAEALKRISLXAIEDIRGLPESERLAVROOR
SKALLTSLHVMKXGNTLKKSRLEGEAFYVLNQDLCYRDDDLGLAEDNNNTABRA
LRVLCGKNSYDLCQILSPRPYAPAGSLYGLKNLKQSDCILFTNSISGILAP"
complement(2039..2782)
/notes="ORF4; similar to Agrobacterium tumefaciens
transposase encoded by GenBank Accession Number AF034769"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAL67336.1"
/db_xref="GI:18265863"
/translation="MRDLIRARAAKRDSSRQRIILSMRLTRDHYVAGHWTGKHT
WLANSPSQSQIATFOHYCQSEQIEDRLQDQIRSLIPWSICNLVQCQALKG
VGLQIATTLVAELGDRFSNPKQLMAFLGVPEYSSGNSIIRPGITKVGNSLRL
LYEAMSVYTPAKVGAWLIYRPDSVTQYSKDIAWKAQQLCSRYSLTAKGKKSQA
ITAVARELTGMWDIALAAGSFSQKQN"
complement(2810..3214)
/notes="ORF5; similar to Agrobacterium tumefaciens
transposase encoded by GenBank Accession Number AF034769"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAL67337.1"
/db_xref="GI:18265864"
/translation="MSENIATAYVGLDVHKETLVAIAAPERLGEVRYGTINNEAQA
VRRLFKQLGYLNIILSYEAGPCGFLYHQLTAMNICKQVIAPSRIPKSPTRKND
HRDAISLARLLRAGELTPVWIKGEFQHTGGRY"
3611..4297
/notes="ORF6; similar to Shigella sonnei transposase
encoded by GenBank Accession Number AB029445"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAL67338.1"
/db_xref="GI:18265865"
/translation="MPLIAPISDDEBLMKAIHKTDKNYAPBLTAMLMHLRGDRVS
DVARTLCARSSVGRINWFTQSGVEGLKSLPAGRARWPFHEHICILLRELVKHSQD
FGYQSRSSWTELLAIKINETITCOLNAGTVRWLPAGIWRRAAPTLRLRDPKDEK
MAAIIHALDECCAEHPVFYDEVDIHLNPKIGADWQLRGQQRKRVTPGQVKPAMWAAT
AKVRRCSAC"
complement(4678..5025)
/notes="ORF7; similar to Escherichia coli L0008 encoded by
GenBank Accession Number AF071034"
/codon_start=1
/transl_table=11
/product="L0008-like protein"
/protein_id="AAL67339.1"
/db_xref="GI:18265866"
/translation="MLPNSRQPTLHRLHLLPLTPYVRCSTAHTEKDNAMLYSLSHOTN
DEGESEWIHTGTGYLLRTDANSYPVLRKLRLGLSKTFRRLVVTLLRCYGVGLHLAA
SAGCLPGLPIFDW"
complement(4892..5116)
/notes="ORF8; similar to Escherichia coli L0008 encoded by
GenBank Accession Number AF071034"

```



```
/codon_start=1
/transl_table=11
/product="L0008-like protein"
/protein_id="AAL67340.1"
/db_xref="GI:18265867"
/transl_table="MGRVVDVVDGIDPAELNNKVAENGYSLSRVVDAPEQQTDSPPPF
TALTDIESLQYRTHGKQRLAVLTVTSDS"
5289..6755
CDS
/notes="ORF9; similar to Escherichia coli R6 protein
encoded by GenBank Accession Number AF081285"
/codon_start=1
/transl_table=11
/product="R6-like protein"
/protein_id="AAL67341.1"
/db_xref="GI:18265868"
/transl_table="MTQKKGITQHISAMKAGISVRSGRRIKESWAKNSVRHWTRK
DPLEAVWDSMLVPLKRPALPTTILEMLQKYPCQYNSLRRTWQRVRWKLQYV
AGEQVMPROHQPLGLSDFTTELKGVVTVIAGLAKHLVHFRLEWHSWVRVLG
GESFSALEGLQALQGLGVPEVHKTDLSLRAMKQGGDGRRELTERTVAALCOHVGX
QGVHNNAGRHENGVSASHLKRRCQALLRGSNDSFTIEEQAQITQVVMHNR
NQDLVKEERLHKPLPERSADYIDELTVXSRSTINVKHYVSVFRLVGLQLLRV
LMDRLUSCYGVSSEWMSCPKRPKERTARRIDFRHVIDSLAKKPGAFCHATLRDI
LPDEWRRLWRILCNLEPDMAGRLMHALKLAAGYDDISVVAKGMSQMLNTPGNVDI
HRLMFLGIXEKALPVNVKQHLSSYEQLLRKGGSQ"
complement(7582..7959)
/notes="ORF10; similar to SwissProt Accession Number
P76365"
/codon_start=1
/transl_table=11
/product="intergenic-region protein"
/protein_id="AAL67342.1"
/db_xref="GI:18265869"
/transl_table="MKTLPDTHVREASRCPSPVTWOTLLTOLLDOHGLTLNDTPPA
DERVIOHIEAGISLCDVNFVLEKVALVTPQPGFSACTRSQSLNSIDILRRARTG
LMTRSYRMYNDIIRGEHSEAK"
complement(8006..8488)
/notes="ORF11; similar to SwissProt Accession Number
P76364"
/codon_start=1
/transl_table=11
/product="intergenic-region protein"
/protein_id="AAL67343.1"
/db_xref="GI:18265870"
/transl_table="MPPFTSMHLPSPKPVSTPADLSSLNYSHPICEDFIVSDTLTET
NYDDNDRPWWGLPCTVTPCFGARLVQEGNQLYLADRAGIRGOFSDADYHLDQAF
FLMKQLEMLTIGELNPRHQHTVLYARGLTCEADTLGSCGYVYMAVYPTLAPATTS"
complement(9093..9314)
/notes="ORF12; similar to SwissProt Accession Number
P76363"
/codon_start=1
/transl_table=11
/product="intergenic-region protein"
/protein_id="AAL67344.1"
/db_xref="GI:18265871"
/transl_table="MKIITRGEAMRIHQOHPASRLFPCTGKYRWHGSTETVYTGREVQ
DIPGLVAFERKDSFGPYVRLMSVTLN"
complement(9377..9868)
/notes="ORF13; similar to SwissProt Accession Number
P76362"
/codon_start=1
/transl_table=11
/product="intergenic-region protein"
/protein_id="AAL67345.1"
/db_xref="GI:18265872"
/transl_table="WEACTWQOLSFLPGEMTTRSLIRALKTLDRHLHPGVAFTS
TRAARWLLINWAGLEREFVLYNNQNLQIATGLTFTGNINTEVEHPREVIKRALY
HNAAVVLAAHNPGEVTPSKADRLITERLQALGLVDIRVPDHLIVGGSQVFSFAEH
GLL"
9389..10348
CDS
/notes="ORF15"
/codon_start=1
/transl_table=11
```

```
/product="unknown"
/protein_id="AAL67347.1"
/db_xref="GI:18265874"
/transl_table="MFGKGNLAADTYQMRHPDIHQACLYQTRFRKPVCLAGCDFT
GRVIMQCHGCGGIVIOGAFNHFPMDFRAVDGAGEEGTGNQLIIVIOIHPELFTLQ
SRVIMQPPCTGTCGGHAGFMKMAVQFGQFQNEALSGCFSGQKGLKPCCLH
SVDDAHNGAFAFRMLQGVIPQTVIMNRHGTLETAGVMTVRHQTGNAGGFAHFAVT"
Query Match 49.0%; Score 1025; DB 1; Length 71684;
Best Local Similarity 99.5%; Pred. No. 4,9e-297;
Matches 1028; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 ATGCGAATAACCACTCTGGCTTCCGTAGTCATTCCTGTCCTCGATTTTCACCCACGACG 60
Db ATGCGAATAACCACTCTGGCTTCCGTAGTCATTCCTGTCCTCGATTTTCACCCACGACG 34034
QY 61 ATAGCTCTCCACAGAGATGATGATGTCCTGGGATCCGGCTATGAGAAAAGTGACT 120
Db ATAGCTCTCCACAGAGATGATGATGTCCTGGGATCCGGCTATGAGAAAAGTGACT 34153
QY 121 AACGACGCCCGAGTGTTCCTGTGATTATCCAGGAGGAATTCAGTCCAGCAGTACCAC 180
Db AACGACGCCCGAGTGTTCCTGTGATTATCCAGGAGGAATTCAGTCCAGCAGTACCAC 34213
QY 181 GATCTGCGGAGGCTCTGAGATCAGTAGAGGTGTGATGTTGAAAGTGGTACGGGTAAA 240
Db GATCTGCGGAGGCTCTGAGATCAGTAGAGGTGTGATGTTGAAAGTGGTACGGGTAAA 34273
QY 241 ACCGAGGGCTGGAATCAGCATCCAGGAATGCCAGCAGTATACACGCTGATCTGATT 300
Db ACCGAGGGCTGGAATCAGCATCCAGGAATGCCAGCAGTATACACGCTGATCTGATT 34333
QY 301 GATGGTCTTCGTCAGCGCGGACGATGACGTGACTCCCAACGGTTTCTGCCATGAT 360
Db GATGGTCTTCGTCAGCGCGGACGATGACGTGACTCCCAACGGTTTCTGCCATGAT 34393
QY 361 ACCGGTTTATGCCCTCTGCGCGCATTTAGCGTATTAGGTTATCAGGGGCGCGATG 420
Db ACCGGTTTATGCCCTCTGCGCGCATTTAGCGTATTAGGTTATCAGGGGCGCGATG 34453
QY 421 TCCACACTGTATGGCTCTGATGCGGTGGCGGTGGTGAATATCATACAGAAAGAT 480
Db TCCACACTGTATGGCTCTGATGCGGTGGCGGTGGTGAATATCATACAGAAAGAT 34513
QY 481 GCAGACAAATGGCTCTCTTCCTGCAATGCAGGGCTGAATCTGCAGGAAAGCAACAAATGG 540
Db GCAGACAAATGGCTCTCTTCCTGCAATGCAGGGCTGAATCTGCAGGAAAGCAACAAATGG 34573
QY 541 GSTAACAGACGACGCTTAAATTTCTGGACAGTGGTCCCTTGTGGATGATTTCTGCAG 600
Db GSTAACAGACGACGCTTAAATTTCTGGACAGTGGTCCCTTGTGGATGATTTCTGCAG 34633
QY 601 CTGCAAGTACGGGTAGCACACACAGCGTCAAGGGTTTCATCGGTCACTCATGAGCGAT 660
Db CTGCAAGTACGGGTAGCACACACAGCGTCAAGGGTTTCATCGGTCACTCATGAGCGAT 34693
QY 661 ACAGACGACGCGGTATTCCTTATCCACGAGTACAGAAATTAATCTTGGTGCAGT 720
Db ACAGACGACGCGGTATTCCTTATCCACGAGTACAGAAATTAATCTTGGTGCAGT 34753
QY 721 CTTGACTGGAAGCGCTCGGAGCAGGATGTGCTCTCGTTTGATATGATACACCCCGGAG 780
Db CTTGACTGGAAGCGCTCGGAGCAGGATGTGCTCTCGTTTGATATGATACACCCCGGAG 34813
QY 781 CGTTATGATAACCGGATGGCAACTGGGGAGTCTGACGGGGGATATGACCGGACCTG 840
Db CGTTATGATAACCGGATGGCAACTGGGGAGTCTGACGGGGGATATGACCGGACCTG 34873
QY 841 CGCTATGACGAAAACAAAATTTTCAGTGGCTATGATCATCTTTTCCCTTCGGAACATGG 900
Db CGCTATGACGAAAACAAAATTTTCAGTGGCTATGATCATCTTTTCCCTTCGGAACATGG 34933
QY 901 AATCGTATCTGAACTGGAAACAGACAGAAAATAAAGTCTGTGAGCTTGTAGCGAGTGTA 960
```


Db 34934 AAATCGTATCTGAACATGGACGAGACAGAAAATAAAGGTCGTGAGCTCGTACGCAGTGTA 34993

QY 961 CTGAAGCGGACAAAATGGGGGCTTGC CGGTCAAGCCGCGGAGCTTAAGGAATCGAACCTT 1020

Db 34994 CTGAAGCGGACAAAATGGGGGCTTGC CGGTCAAGCCGCGGAGCTTAAGGAATCGAACCTT 35053

QY 1021 ATCTCGAATTCAT 1033

Db 35054 ATCCTGAATTCCT 35066

RESULT 14

AE013687

LOCUS 10029 bp DNA linear BCT 26-JUL-2002

DEFINITION *Yersinia pestis* KIM section 87 of 415 of the complete genome.

ACCESSION AE013687 AE009952

VERSION AE013687.1 GI:21957573

KEYWORDS

SOURCE *Yersinia pestis* KIM

ORGANISM *Yersinia pestis* KIM

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Yersinia.

REFERENCE 1 (bases 1 to 10029)

AUTHORS Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F., Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C., Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V., Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S., Blattner, F. R. and Perry, R. D.

TITLE Genome Sequence of *Yersinia pestis* KIM

JOURNAL J. Bacteriol. 184 (16), 4601-4611 (2002)

PUBMED 12142430

REFERENCE 2 (bases 1 to 10029)

AUTHORS Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F., Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C., Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V., Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S., Blattner, F. R. and Perry, R. D.

TITLE Direct Submission

JOURNAL Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES

source

1..10029

Location/Qualifiers

/organism="Yersinia pestis KIM"

/mol_type="genomic DNA"

/strain="KIM"

/db_xref="taxon:187410"

complement(91..1089)

/gene="y0843"

complement(91..1089)

/gene="y0843"

/function="regulator"

/note="similar to sorbitol operon regulator; residues 14 to 327 of 332 are 68.47 pct identical to residues 1 to 314 of 317 from GenPept: >gb|AA03730.1| (AE006201) unknown [Pasteurella multocida]"

/codon_start=1

/transl_table=11

/product="putative transcriptional regulator"

/protein_id="RAM84428.1"

/db_xref="GI:21957574"

/translation="MISGRGQLSMEKQIPQDWELLTEIAVSYQDMEQTEIAHKFGISRKVGALLAKRATGEIVNRYHPVFTRELQDHFHGLKRALIALDYQDEDEQIRQVAALSRKATGELKDVNTAVGGRNVAADHSVVPFRHCKPFCIGGTHRPGVADNADHSRLRAKFKGTSITLAPAVENRALKEAFMGNTIKETLDRKADWALVIGIGMNSYVYELGWFTPQEIIVDAISLHQVTGEIAGYDFDAQGPANTVMNDVIGISTQELROI PCVIAIAAENTKALLCALRTGATIDILATSASNARTILKMIQHS"

gene

complement(1617..2906)

/gene="arsB"

/note="synonym: y0844"

complement(1617..2906)

/gene="arsB"

/function="transport; drug/analogue sensitivity"

gene

complement(1617..2906)

/gene="arsB"

/function="transport; drug/analogue sensitivity"

QY	1186	ACGATGACATTGGCTGACTGCGGAGCGCTGATGAACATCATGACCAATTCGGGGA	1245
Db	8746	CTGGACTCACTGGCATTAAACCTATGGCGCAGCTTATGAATATCAATACCTTTGGCGGC	8805
QY	1246	CACATCAGTCGCGCTGATATCTGGTCTGGGATGTCGATGCTGGAGCGCTGAAGGC	1305
Db	8806	CATATTAGCCACAGAGCTATCTGGTGGGACCCACAGATTAACCTGGACGGTGAAGGG	8865
QY	1306	GGTGTGACCGGATATAGGACCCAGAGTGGGAGCTGATCAATAAGGCGATTAGTGGT	1365
Db	8866	GGTGTGACCGGATATAGGACCCAGAGTGGGAGCTGATCAATAAGGCGATTAGTGGT	8925
QY	1366	GTGTCCGGGAGGAAACAAATCTACTTGGTAAACCCCGACCTGAAGCCGGAAGAGC	1425
Db	8926	GTCAAGGCCAAGCACCATCACGACGATTGGTAACCTTAAGCTTGAACCCAGAGTAGT	8985
QY	1426	GTCAAGTTATGAGTGGGCTGATTACGATAACCCCGCGTCTGAATGCCATGTCA	1485
Db	8986	GTCAACACCGAGTGGCTGCTATTATTAACAAACGAGGATTTGGCGTAAATGTGACG	9045
QY	1486	GGTTTATGACTGCTTCTCCAAAGAAATGTCTCTTATTCATAATGATACCAAT	1545
Db	9046	CTGTTCACACCGCTTCAGGAACAAATCAACAGCGTTTCGATTGATAATACCCAGC	9105
QY	1546	AGCTATGTAAACAGCGGAGGCGGCTGACGGTGTGAATTTCCGCGCATTGCGG	1605
Db	9106	ACCTACACCGTGGTAAGCGATCACGCGGTTATGAGTGCATCCACCATPACA	9165
QY	1606	CTGTGCTCAGAGGATGTCAGCTGTCTCACTGAATTAACCTTGGACCCGAAAGTGAACAGT	1665
Db	9166	CTGTGCTCAGAGGATGTCAGCTGTCTCACTGAATTAACCTTGGACCCGAAAGTGAACAGT	9225
QY	1666	GATGCTGATTAACAGTGGCGGCTGAGTTATACCCCTGACACATGTTGAATGCGAA	1725
Db	9226	GATGGCAACATAAAGGGCTCGCTGACGAATATACACCGAAATATGTTTAAATGCCCG	9285
QY	1726	CTGAACCTGGCAGATCACCGAAGGTTGGCATCATGCTGGTCCCGTTATCGCGGAA	1785
Db	9286	CTGAACCTGGATATCAACGAGCAGTTGTCCACTTGTGTAAGCGAGTATGCGAGTAA	9345
QY	1786	ACACCGCTTCCACGAGATTAATCGTCACTGAGCGCTGTACAGAAAGATGTATGAT	1845
Db	9346	ACGCGCAGTTTACCCCAAAATTAACGGCAATTTGTCTGCGGCAACAAAGTGTGTACAC	9405
QY	1846	GAGAAAGGAGATACCTGAAAGCTGACGGTGGTGGATCGAGTCTGCTGGAGATG	1905
Db	9406	AATTTAGCTCAGAGTTTAAACCTTTAGCGTAAATTTGGGGTGGCTTATAGGTC	9465
QY	1906	ACGATGCCCTGACGCTGAAATGCTGCGGTGAATAACCTCTCAACAAAGATTAACAGTAC	1965
Db	9466	ACTAAGGATGTGACACTGAATGTGCGAGTCAATAACCTTGGATAAAGATTTACTCGC	9525
QY	1966	GTGAGCTGTACAGTCCGCTAAGATAGCTGTATGCGGTTGATTTCTTCAGACGGA	2025
Db	9526	ACCCACATTTTCTGTGGGAGACGCACTACTGCGGGTGAATTTACTTCACTCCGCT	9585
QY	2026	TCATCAACACAGGATATGTATCTGACCTGAGCAAAATTAATCTGATGCTGCTGAATCAT	2085
Db	9586	CRAAGACTGCGGTTATGTGTACTCTGCGCGTAAATTAATCTGCGTATCGGTGATTAAT	9645
QY	2086	TTCT 2089	
Db	9646	TTCT 9649	

RESULT 15

AJ414156/C
 LOCUS 220050 bp DNA linear BCT 21-NOV-2002
 DEFINITION Yersinia pestis strain CO92 complete genome; segment 16/20.
 ACCESSION AJ414156 AJ509842
 VERSION AJ414156.1 GI:15981150
 KEYWORDS

SOURCE	Yersinia pestis CO92
ORGANISM	Yersinia pestis CO92
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Yersinia.
AUTHORS	1 (bases 1 to 220050) Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G., Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L., Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdano-Tarraga, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Feltwell, T., Hanlin, N., Holroyd, S., Jagels, K., Leather, S., Karyshev, A.V., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrall, B.G. Genome sequence of Yersinia pestis, the causative agent of plague Nature 413 (6855), 523-527 (2001) 21470413 PUBMED 11586360
TITLE	2 (bases 1 to 220050)
JOURNAL	Parkhill, J.
MEDLINE	Direct Submission
PUBMED	Submitted (04-OCT-2001) Submitted on behalf of the Yersinia sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT	Details of Y. pestis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/Y_pestis/).
FEATURES	Location/Qualifiers 1. 220050 /organism="Yersinia pestis CO92" /mol_type="genomic DNA" /strain="CO92" /db_xref="taxon:214092" /note="biovar: Orientalis" complement(191..715) /gene="arol" /note="synonym: YPO3215" complement(191..715) /gene="arol" /note="synonym: YPO3215" complement(191..715) /EC_number="2.7.1.71" /note="Similar to Escherichia coli shikimate kinase II AroL SW:AROL_ECOLI (P08329) (174 aa) fasta scores: E(): 1.5e-31, 55.4% id in 168 aa, and to Erwinia chrysanthemi shikimate kinase AroL or AroM SW:AROL_ERWCH (P10880) (173 aa) fasta scores: E(): 0, 63.7% id in 168 aa" /codon_start=1 /transl_table=11 /product="shikimate kinase II" /protein_id="CAC92450.1" /db_xref="GI:15981151" /db_xref="GOA:Q8ZC15" /db_xref="SWISS-PROT:Q8ZC15" /translation="MTQTFMGARGAGKTTIGKALQAQALGYEVVDLDMQOTSMTVASVSESGDGRFLRESMALQVATPVTGAGVLSSENRAFMRDHGRVYIDRA SAVALARRAEDDEAQRSLTGPIVEILDVLASREALYQDVAAHVLDGTQTPSLVVEQILQMLTGEVMVK" complement(209..709) /gene="arol" /note="Pfam match to entry PF01202 SKI, Shikimate kinase, score 245.00, E-value 1.1e-69" complement(470..544) /gene="arol" /note="PS01128 Shikimate kinase signature." complement(668..691) /gene="arol" /note="PS00017 ATP/GTP-binding site motif A (P-loop)." complement(962..1519) /gene="YPO3216" complement(962..1519) /gene="YPO3216" /note="Similar to Rhizobium meliloti phospholipid N-methyltransferase PmtA TR:AAG10237 (EMBL:AF201699) (200 aa) fasta scores: E(): 1.1e-07, 25.9% id in 185 aa, and to Bradyrhizobium japonicum PmtA protein TR:Q9LCT2
misc_feature	
misc_feature	
misc_feature	
gene	
CDS	

QY 358 AATACCGGTTTCATGCCCCCTCTCGCGCCATTGAGCGTATTGAGGTTATCAGGGGGCCG 417
 Db 147953 AATACCGGTTTCATGCCCCCTCTCGCGCCATTGAGCGTATTGAGGTTATCAGGGGGCCG 147894
 QY 418 ATGTCACACCTGTATGGCTGTGATCGATGGCGGTGTGATGATATCATTTACAGAAAG 477
 Db 147993 ATGTCACACCTGTATGGCTGTGATCGATGGCGGTGTGATGATATCATTTACAGAAAG 147834
 QY 478 AATGACAGCAAAATGGCTCTCTTCGGTCAATGCGAGGCTGAATCTGACAGAAAGCAACAA 537
 Db 147833 ATCACCAGAGGCTGGCCAGCTCAGCAACGCTTGAAACACACTTCAGAGAAATACGGCT 147774
 QY 538 TGGGTTAACAGCAGCCAGTTTAATTTCTGAGCAGTGGTCCCTTGTGGATGATTTCTGTC 597
 Db 147773 TTTGGTACACAGTTCAAAAGTTCTCTTTTACTCCAGTGGCGCAGCGTTGAAGATCAACTG 147714
 QY 598 AGCTGACAGTTACCGGTAGCACACACAGCGTCAGGTTTCATCGTFCATC-----A 651
 Db 147713 GGGCTGCTGTACGTGCACAGATTTCCGGCGGATGCCCTCTCGGTTGAATCCTCCAAT 147654
 QY 652 CTGAGCGATACAGCAGCCAGCTTATCTTATCCACGAGTACAGAAATTAATCTT 711
 Db 147653 ACAGGAGTAGAATTAAGCCGCTGCTGGTCCCAATCCAGTAAAGCAGACATATAATCTC 147594
 QY 712 GGTGACGCTCTGACTGGAAGGCTCGAGCAGGATGTGCTCTGTTTGTATGATGATACC 771
 Db 147593 GGTGGCAACTATTTTGGCAATTTGAATACCCAAAGCAGCTTGTGTTGGATGGTGATTT 147534
 QY 772 ACCGGCAGCGTTATGATACCGGGATGGCAACTGGGA-----GTCTGACGGGGGA 825
 Db 147533 GCTAACCAAAATATGATACAGTGCACCACTAGGTACCATTTGTGTCAGGCGGT 147474
 QY 826 ATGACCGGACCTTGCGCTATGAGCAACAAATTTAGCTGGCTATGATCATCTTTC 885
 Db 147473 TAGCAAGATCTCTGCGTTATCAACCGGTAATAAATTAACCTTGGGCAACGATATCGTTTG 147414
 QY 886 ACCTTCGGAACATGGAATCGTATCTGAACTGGAACGAGACAGAAATTAAGTCTGTAG 945
 Db 147413 GACTTTGGTACCTTGGATTCGACCTTATCTTATACCAACTGAAATATTGGCCG---C 147357
 QY 946 CTGTGACGAGTGTACTGAAAGCGCAAAATGGGGCTTTCGGGTGAGCGCGGAGCTT 1005
 Db 147356 TTAATCACCAATCATCCGTCGCGCCCAAGGCTCTGGGTGGCAGGTGAAAGAGCTCTGTG 147297
 QY 1006 AAGCAATCGAACCTTATCTGTAATTCATTTACTGCTTACCCCTCTGGGGAATCTCATCTG 1065
 Db 147296 AAAATACCAATATCATCTTGTATCAAGCTGCTCGCCCGCTGGGTGATTAACACATG 147237
 QY 1066 GTTACGGTGGGGCGAGTTTCAGACTCGTCCATGAAAGCAGGATTTCTTCCGACG 1125
 Db 147236 GTGACCCCTTGGCGTCAATATTGGAATGCCATCATGAAAGCGGCATCTACTGGCAAT 147177
 QY 1126 ACAGGTGAACCTTTCGGGAGAAAGCTGTGCGGTTATTTGCTGAGGATGAGTGCATCTC 1185
 Db 147176 AATGGCGTAGGTTTGCAAGAATGCTGTGCTCATTTATCAGGGAAGATGAATGGCGTCTG 147117
 QY 1186 ACGGATGCACTTTCGGTGTGCTGGGCGAGCGCTATGAAACATCATGCAATTCGGGGGA 1245
 Db 147116 CTGCACTACTGSCATTAACTATGCGCACTGATGATGATATCAACTACGTTTGGCGG 147057
 QY 1246 CACTTCAGTCCGGTGCATATCTGCTGCGATGTGGGATGCGGATGCGCTGAGCGTGAAGGC 1305
 Db 147056 CATATTAGCCCGAGCCCTATCTGTTGTGGGCGCACAGGATAACTGGACGCTGAAAGGG 146997
 QY 1306 GGTGTACCAACGGGATATAGGCACCCAGATGGGCGAGCTACATAAAGGGATTTAGTGT 1365
 Db 146996 GGTGTACGACGCGGGTATTAACCCCACTCTAGCCAGTTGCATATGCGATTTAGCGGT 146937
 QY 1366 GTGTCGGGCGAGGAAAAACAATCTTCTTGTAAACCCGAGCTGAACCGGAGAGAGC 1425
 Db 146936 GTCACAGGCCAAGGCCCATCAGACGATTGGTAACCCCTAAGCTTGAACCCAGAGTAGT 146877

QY 1426 GTCACTTATGAGCTGGGTGTATTACGATAACCCCGCGCTGCAATGCAATGCAATGCA 1485
 Db 146876 GTCAACACCGAGTGGGTGTCTATTATGAAACGAAACGGGATTTGGCGCTAATGTGACG 146817
 QY 1486 GGTATTATGACTGACTTCTCCAAAGAGTTGTCTCTTATTTCCATTAATGATTAACCAAT 1545
 Db 146816 CTGTTCCAAACCGCTTCAGAAACAAATCAACACGCTTTCGATTGATATACCCAGC 146757
 QY 1546 AGCTATGTAACACAGCGGAAAGCCCGCTTGCACGCTGTGGAATTTGCGGACATTCGCG 1605
 Db 146756 ACCTACACCAACGCTGGTAAACGCTACGTTAGGTTAGGTCGATCCACCATACCA 146697
 QY 1606 CTGTGCTCAGAGATGTCCAGCTGTCTGATTAACCTGGAACCCGAGTGAACACAGT 1665
 Db 146696 CTGTGCTCAGAGTGTGATGCTGGGAATAAATACACCTTTACCGGATGAGTGAACAGAAA 146637
 QY 1666 GATGTGATAACAAAGGTGCGCGCTGAGTTATACCCCTGAACACATGGTGAATGCGAAA 1725
 Db 146636 GATGCAACAATAAAGGGCTCGCTGACGAATACCGAAATAATATGTTTATGCCCCG 146577
 QY 1726 CTGAACCTGGCAGATCACCGAAGAGTGGCATCATGCTGGGTGGTCCCGTTATCGCGGAAA 1785
 Db 146576 CTGAACCTGGAATATCAACGAGCAGTTGTCCACTTTGGTTGAAAGCGGAGTATCGCAGTAA 146517
 QY 1786 ACACACGTTTTCACCGCAGATTTATGCTGCTGAGCGCTGTACAGAAAGTGTATGAT 1845
 Db 146516 ACGGACAGTTTATCCCAAAATTTACCGAAATTTGCTCGGCAACAAAGTGGTTTACAAC 146457
 QY 1846 GAGAAAGGAGAAATACCTGAAAGCCTGGACGCTGGTGGATGAGTCTGCTGGAAGATG 1905
 Db 146456 AATTTAGGCTCAGAGTTTAAACCTTTTAGCGTACTTAATTTGGGGTGGCTTAAAGTC 146397
 QY 1906 ACGGATGCGCTGAGCTGATGCTGGGTGAATAACCTGCTCAACAGGATTAACAGTAC 1965
 Db 146396 ACTAAGGATGTGACACTGAATGCTGAGTCAATAACCTGTTGGATAAAGATTTTACTCGC 146337
 QY 1966 GTGAGCCTGTACAGTGGCGGTGAAGATGAGTGTGATGCTGCGGTGATTTCTCCAGCGGA 2025
 Db 146336 ACCACATTTTGTGCTGGGGAACGGCACCTACTTGGGGTGAATTTTCACTCTGCT 146277
 QY 2026 TCATCAACACAGGATATGTATACCTGAGCGGAAATTTACTGGATGCTGCTGAACATACAG 2085
 Db 146276 CAAAGCACTGCGCGTTATGTTGTACCTGGCGGTAATTTACTGGGTATCCGTTGAATGTAAT 146217
 QY 2086 TTCT 2089
 Db 146216 TTCT 146213

Search completed: October 12, 2004, 15:56:10
 Job time : 8126 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: October 13, 2004, 11:41:17 ; Search time 131 Seconds
(without alignments)

10072.497 Million cell updates/sec

Title: US-10-625-972-4

Perfect score: 3732

Sequence: 1 atcggaataaccactctggc.....cgtgaactatcagttctga 2091

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DB=frame+ n2p.model -DEV=xlh
-O/cn2_1/USPTO spoal/US10625972/runat 12102004 130105 26819/app query.fasta_1.2247
-DB=SPREMBL_25 -QFWT=fasan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40 cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10625972 @CGN 1 1 177 @runat 12102004 130105 26819 -NCPU=6 -CPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3632	97.3	696	16 Q9LAP1	Q9lap1 escherichia

RESULT 1

ID	Q9LAP1	PRELIMINARY;	PRT;	696 AA.
AC	Q9LAP1	Q8X9P7;		
DT	01-OCT-2000	(TRENBLrel. 15, Created)		
DT	01-JUN-2002	(TRENBLrel. 21, Last sequence update)		
DT	01-JUN-2003	(TRENBLrel. 24, Last annotation update)		
DE	Adhesin (Putative receptor) (Iha adhesin).			
GN	IHA OR Z1178 OR Z1617 OR ECS1360.			
OS	Escherichia coli O157:H7.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=83334;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=86-24;			
RX	MEDLINE=20143752; PubMed=10678953;			
RA	Tarr P.I., Bilge S.S., Vary J.C. Jr., Jelacic S., Habeeb R.L.,			
RA	Ward T.R., Baylor M.R., Besser T.E.;			
RT	"Iha: a novel Escherichia coli O157:H7 adherence-conferring molecule			
RT	encoded on a recently acquired chromosomal island of conserved			
RT	structure."			
RL	Infect. Immun. 68:1400-1407(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			

ALIGNMENTS

Q8fdx2 escherichia
Q93se1 escherichia
Q87518 escherichia
Q8192 yersinia pe
Q8v68 escherichia
P0847 bordetella
Q9pg3 campylobact
Q7vi87 helicobacte
O07651 campylobact
Q87g14 yersinia pe
Q87lk9 vibrio para
Q93rd3 agrobacteri
Q8v48 escherichia
Q8fa7 escherichia
Q93b0 escherichia
Q88x7 shewanelle
Q8896 pseudomonas
Q82597 salmonella
Q82n0 salmonella
Q8x63 escherichia
Q8cv8 escherichia
Q91212 pseudomonas
Q8404 acinetobact
Q88kp8 pseudomonas
Q7vv12 bordetella
Q7w19 bordetella
Q7w108 bordetella
Q8cwal escherichia
Q45341 bordetella
Q8xbw7 escherichia
Q821b8 shigella fl
Q9r19 escherichia
Q8cy02 escherichia
Q8pj6 xanthomonas
Q93x7 escherichia
Q82818 salmonella
Q68598 pseudomonas
Q91527 pseudomonas
Q8gh18 escherichia
Q8z40 salmonella
Q8z4h1 salmonella
Q8rm40 salmonella
Q8zmn0 salmonella

RC STRAIN=86-24;
RA Tarr P.I., Bilge S.S., Vary J.C., Jelacic S., Habeeb R.L., Ward T.R.,
RA Baylor M.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouisis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 050952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genome comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
DR EMBL: AF126104; AAF36432.2; -;
DR EMBL: AE005310; AAG55732.1; -;
DR EMBL: AE005274; AAG55323.1; -;
DR EMBL: AP002554; BAB34783.1; -;
DR PIR: G85607; G85607.
DR PIR: H90798; H90798.
DR HSP: P05825; IFEF.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0004872; F:receptor activity; IEA.
DR GO: GO:0005215; F:transporter activity; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR000531; TonB boxC.
DR Pfam: PF00593; TonB dep Rec; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 696 AA; 76512 MW; 7651E329169D9064 CRC64;

Alignment Scores:
Pred. No.: 3,18e-223 Length: 696
Score: 3632.00 Matches: 695
Percent Similarity: 99.86% Conservative: 0
Best Local Similarity: 99.86% Mismatches: 1
Query Match: 97.32% Indels: 0
DB: 16 Gaps: 0

US-10-625-972-4 (1-2091) x Q9LAP1 (1-696)

QY 1 ATCCGATACCACTCTGCTCCGTCAGTATTCCTGTCCTCGGATTTTCACCCGAGC 60
Db 1 MetArgilethrThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSer 20
QY 61 ATAGCTGCTCGAGGATGTGATGTTGTCGCGATCCGGCTATCAGAAAAAGCTGACT 120
Db 21 IleAlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysIleLeuThr 40
QY 121 AACGAGCCGCCAGTCTTTCTGTGATTACCCAGGAGGAATTCAGTCACGCCAGTACCAC 180
Db 41 AsnAlaAlaAlaSerValSerValIleSerGlnGluGluLeuGlnSerSerGlnTyrHis 60
QY 181 GATCTCGCGAGGCTCTGAGATCAGTAGAGGCTGTGGATGTTGAAAGTGGTACGGGTAAA 240
Db 61 AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys 80
QY 241 ACCGAGGCTCGAATACAGTCCAGGAATGCCAGGAGTACACGCTGATCTGATT 300
Db 81 ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyrThrLeuIleLeuIle 100

QY 301 GATGTGTTCTGTCAGGCGGAGCAGTGACGTGACTCCCAACGGTTTTTCTGCCATGAAT 360
Db 101 AspGlyValArgGlnGlyGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
QY 361 ACCGGTTTCATGCCCCCTCTGGCGCCATTCAGCGTATTGAGGTATTACAGGGGCCGATG 420
Db 121 ThrGlyPheMetProProLeuAlaAlaIleGluArgIleGluValIleArgGlyProMet 140
QY 421 TCCACACTGATGCTCTGATGCGATGGCGCGCTCTGGTGAATATCATACCAGAAAGAAT 480
Db 141 SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAsnIleIleThrArgLysAsn 160
QY 481 GCAGCAAAATGGCTCTCTTCGCTCAATGCAGCGCTGAATCTGCAGGAAAGCAACAATGG 540
Db 161 AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp 180
QY 541 GGTAAACAGCAGCCAGTTTAATTTCTGGAGCAGTGTCCCTCTGTGGATGATTTCTGTCCAGC 600
Db 181 GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspSerValSer 200
QY 601 CTGAGGTACCGGTAGCACACACACGCTCAGCGTTTCATCGGTACATCAGTACGAT 660
Db 201 LeuGlnValArgGlySerThrGlnGlnArgGlnGlySerSerValThrSerLeuSerAsp 220
QY 661 ACAGCGGCGCGCTATTCCTTATCCACGAGTGCACAGATTAATATCTGTGTGACGCT 720
Db 221 ThrAlaAlaThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg 240
QY 721 CTTGACTCGAAGCGCTCGGAGCAGATGTGCTCTGTTTTCATATGATATGATACCCCGCAG 780
Db 241 LeuAspTrpLysAlaSerGluGlnAspValLeuTrpPheAspMetAspThrThrArgGln 260
QY 781 CTTTATCATACCGGATGGCAACTCGGAGTCTGACGGGGGATATGACCGAACCTG 840
Db 261 ArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGlyTyrAspArgThrLeu 280
QY 841 CGCTATGAGCAAAACAAATTTTCAGCTGCTGATCATCATCTTTCACCTTCGGAACATGG 900
Db 281 ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspHisThrPheThrPheGlyThrTrp 300
QY 901 AAATCGTATCTGAACTGGAACGAGACAGAAAAATAAGGTCTGAGCTGTGTACCGAGTGA 960
Db 301 LysSerTyrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal 320
QY 961 CTGAAGCGCGACAAATGCGGCTTCGCGTCAGCCGCGGAGCTTAAGGAATCGAACCTT 1020
Db 321 LeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu 340
QY 1021 ATCTGAAATTCATTACTGCTTACCCCTCTGGGAGAATCTCATCTGTTACGGTGGGGGC 1080
Db 341 IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly 360
QY 1081 GAGTTTCAGAGCTCGTCCATGAAGCGAGTGTCTCTTCCAGCAGCAGGTGAACCTTTC 1140
Db 361 GluPheGlnSerSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe 380
QY 1141 CGGCAAAAAAGCTCGTGGTATTGTTGAGGATGAGTGGCATCTCAGGATGACCTTGGC 1200
Db 381 ArgGlnLysSerTrpSerValPheAlaGluAspGluTyrHisLeuThrAspAlaLeuAla 400
QY 1201 CTGACTCGGGGAGCGCTATGAACATCATGACCAATTCGGGGGACACTTCAGTCGCGT 1260
Db 401 LeuThrAlaGlySerArgTyrGluHisHisGluGlnPheGlyGlyHisPheSerProArg 420
QY 1261 GCATATCTGCTCGGATGTGGCAGATGCTGACGCTGAAAGCGGTGTGACCAACGGA 1320
Db 421 AlaTyrLeuValTrpAspValAlaAspAlaTrpThrLeuLysGlyGlyValThrThrGly 440
QY 1321 TATAAGGCCCCAGAAATGGGCGAGCTTACATAAAGGATTAGTGTGTGTCGGGCGAGGA 1380
Db 441 TyrLysAlaProArgMetGlyGlnLeuHisLysGlyIleSerGlyValSerGlyGlnGly 460
QY 1381 AAAACAAATCTACTTGTGTAACCCCGACCTGAAGCCGGAAGAGAGCGTCACTTATGAGCT 1440

Db 461 LysThrAsnLeuLeuGlyAsnProAspLeuLysProGluGluSerValSerTyGluAla 480
 QY 1441 GGGGTGTATAGGATAACCCCGCGCTGTAATGCCAATGTCACAGGTTTATGACTCAC 1500
 Db 481 GlyValTyTrpAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp 500
 QY 1501 TTCTCCAAACAGATTGCTCTTATTCCTAATAATGATAACACCAATAGCTATGTAACAGC 1560
 Db 501 PheSerAsnLysIleValSerTySerIleAsnAspAsnThrAsnSerTyValAsnSer 520
 QY 1561 GGAAGGCGCGGTTCGACGGTGTGAATTTGCCGGCACATTGCCCTGTGTGTCAGAGGAT 1620
 Db 521 GlyLysAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp 540
 QY 1621 CTCACGCTGCTCACTGAATTAACCTGGACCGGACCTGAACACGCTGATGTAACAAA 1680
 Db 541 ValThrLeuSerLeuAsnTyTrpThrArgSerGluGlnArgAspGlyAspAsnLys 560
 QY 1681 GGTGGCGCGCTGAGTTATACCCCTGAACACATGTTGAATGCGAACTGAACCTGGCAGATC 1740
 Db 561 GlyAlaProLeuSerTyThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle 580
 QY 1741 ACCGAGAGCTGGCATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1800
 Db 581 ThrGluGluValAlaSerTrpLeuGlyAlaArgTyArgGlyLysThrProArgPheThr 600
 QY 1801 CAGAATTATTCCTCACTGAGCGCTGTACAGAGAGAAAGTGTATGATGAGAGAGGAGATAC 1860
 Db 601 GlnAsnTySerSerLeuSerAlaValGlnLysLysValTyAspGluLysGlyGluTy 620
 QY 1861 CTGAAAGCTGACGCGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
 Db 621 LeuLysAlaTrpThrValValAlaGlyLeuSerTrpLysMetThrAspAlaLeuThr 640
 QY 1921 CTGAATGCGCGTGAATTAACCTGCTCAACAGGATTACAGTACGCTGAGCGCTGATACAGT 1980
 Db 641 LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTySerAspValSerLeuTySer 660
 QY 1981 GCGGTGTAAGTAGTACGCTGATGCGGTGATTAATTCACAGCGGATCATCAACACAGGA 2040
 Db 661 AlaGlyLysSerThrLeuTyAlaGlyAspTyPheGlnThrGlySerSerThrThrGly 680
 QY 2041 TATGTATACCTGAGCGAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 2088
 Db 681 TyrValIleProGluArgAsnTyTrpMetSerLeuAsnTyGlnPhe 696

RESULT 2

Q8FDX2 PRELIMINARY; PRT; 696 AA.
 AC Q8FDX2;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative receptor.
 GN C3610.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22398234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RA "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli."
 RT Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
 RL EMBL: AE016766; AA02058.1;
 DR GO: GO:0016020; C:membrane; IEA.

DR GO: GO:0004872; F:receptor activity; IEA.
 DR GO: GO:0005215; F:transporter activity; IEA.
 DR GO: GO:0006810; P:transport; IEA.
 DR InterPro: IPR000531; TonB_boxC.
 DR Pfam: PF00593; TonB_dep_Rec; 1.
 KW Receptor; Complete Proteome.
 SQ SEQUENCE 696 AA; 76511 MW; F96D10B59C553CE8 CRC64;

Alignment Scores: 6.62e-223 Length: 696
 Pred. No.: 3627.00 Matches: 694
 Score: 3627.00
 Percent Similarity: 99.86% Conservative: 1
 Best Local Similarity: 99.71% Mismatches: 1
 Query Match: 97.19% Indels: 0
 DB: 16 Gaps: 0

US-10-625-972-4 (1-2091) x Q8FDX2 (1-696)

QY 1 ATCGAATAACCACTCTGGCTTCGCTAGTCATCCCTGCTCGGATTTTCAGCCAGCAGC 60
 Db 1 MetArgIleThrThrLeuAlaSerValIleProCysLeuGlyPheSerAlaSerSer 20
 QY 61 ATAGCTCTCCAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
 Db 21 IleAlaAlaAlaGluAspValMetIleValSerAlaSerGlyTyThrGluLysLysLeuThr 40
 QY 121 AACCCAGCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
 Db 41 AsnAlaAlaAlaSerValSerValIleSerGlnGluLeuGlnSerSerGlnTyHis 60
 QY 181 GATCGCGGAGGCTCTGAGATCAGTAGACGGTGTGATGATGATGATGATGATGATGATGAT 240
 Db 61 AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys 80
 QY 241 ACCGAGGCTGGAATCAGCATCCGAGGATCCAGCCAGTTCACAGCTGATGATGATGAT 300
 Db 81 ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyThrLeuIleLeuIle 100
 QY 301 GATCGTGTCTGCTAGCGCGGAAGCAGTGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
 Db 101 AspGlyValArgGlnGlyGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
 QY 361 ACCGCGTTCATGCCCCCTCTGCGCCCATTCAGCGGATTCAGGTTATCAGGGGCCCATG 420
 Db 121 ThrGlyPheMetProProLeuAlaAlaIleGluArgIleGluValIleArgGlyProMet 140
 QY 421 TCCACACTGTATGCTCTGATGCGATGCGGCTGTGTAATATCATTCACAGAAAGAAAT 480
 Db 141 SerThrLeuTyGlySerAspAlaMetGlyValValAsnIleIleThrArgLysAsn 160
 QY 481 GCAGACAAATGGCTCTCTTCCTCAATCCAGGGCTGATTCGACGAGAAAGCAACAAATGG 540
 Db 161 AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp 180
 QY 541 GGTAAACAGCCAGCTTAATTTCTGGAGCAGTGGTCCCTCTGTTGATGATGATGATGATGAT 600
 Db 181 GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspAspSerValSer 200
 QY 601 CTGACAGGTACCGGTAGCACAAACAGCGTTCAGCGGTTTCATCGGTACATCATCTACGCGAT 660
 Db 201 LeuGlnValArgGlySerThrGlnGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGln 220
 QY 661 ACAGCAGGCACGGTATTCCTTATCCACGAGTACAGGATTCAGGATTCAGGATTCAGGATTC 720
 Db 221 ThrAlaAlaThrArgIleProTyProThrGluSerGlnAsnTyAsnLeuGlyAlaArg 240
 QY 721 CTGACTCGAAGGCTCGGAGCAGATGCTGCTGTTGATGATGATGATGATGATGATGATGATGAT 780
 Db 241 LeuAspTrpLysAlaSerGluGlnAspValLeuTrpPheAspMetAspThrThrArgGln 260
 QY 781 CGTTATGATAACCGGATGGGCAACTGGGAGTCTGACGGGGGATATGACCGGACCCCTG 840
 Db 261 ArgTyAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGlyTyAspArgThrLeu 280

QY 841 CGCTATGAGCGAACAATAATTCAGCTGGCTATGATCATATCTTTACCTTCGGAACATGG 900
 Db 281 ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspHisThrPheThrPheGlyThrTrp 300
 QY 901 AAATCGTATCTGAAGTGGACGACAGCAAAATTAAGTGTGAGCTGTGACCGAGTGA 960
 Db 301 LysSerTyrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal 320
 QY 961 CTGAAGCGCGCAAAATGGGGCTTTCGGGTCAGCGCGGAGCTTAAGGAATCGAACCTT 1020
 Db 321 LeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu 340
 QY 1021 ATCTCGAATTCATCTACTGCTTACCCCTCTGGGAGAACTCATCTGTTACGGTGGGGGC 1080
 Db 341 IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly 360
 QY 1081 GAGTTTCAGAGCTCGTCCATGAAGACGAGTGTCTCTCCAGCAGCAGTGAACCTTTC 1140
 Db 361 GluPheGlnSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe 380
 QY 1141 CGGCAAGAAAGCTGGTGGTATTTGTCAGGATGAGTGGCATCTCAGGATGCACTTGGC 1200
 Db 381 ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspAlaLeuAla 400
 QY 1201 CTGACTCGGGCAGCGCTATGACATCATGACAAATTCGGGGACACTTCAGTCCGCT 1260
 Db 401 LeuThrAlaGlySerArgTyrGluHisGlnPheGlyGlyHisPheSerProArg 420
 QY 1261 GCATATCTGCTGGGATGGCAGATGCTCGACGCTGAAGCGCGTGTGACACGGA 1320
 Db 421 AlaTyrLeuValTrpAspValAlaAspAlaTrpThrLeuLysGlyGlyValThrThrGly 440
 QY 1321 TATAGGACCCAGATGGGGCAGCTACATAAGGATGATAGTGTGTCTGGGCGAGGA 1380
 Db 441 TyrLysAlaProArgMetGlyGlnLeuHisLysGlyLeuSerGlyValSerGlyGlnGly 460
 QY 1381 AAAACAAATCTACTTGTGAACCCGACCTGAAGCCGGAAGAGAGCTGAGTTATGAGCT 1440
 Db 461 LysThrAsnLeuLeuGlyAsnProAsnLeuLysProGluSerValSerTyrGluAla 480
 QY 1441 GGGGTGATTAACAAACCCCGCTGCAATGCCAATGTCAAGTTTATGATGAC 1500
 Db 481 GlyValTyrTyrAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp 500
 QY 1501 TTCTCCAAAGATGTCTTATTCATAAATGATAACCAATAGCTATGTAACACAGC 1560
 Db 501 PheSerAsnLysIleValSerTyrSerIleAsnAspAsnThrAsnSerTyrValAsnSer 520
 QY 1561 GGAAGCGCGGTGACCGGTGGAAATTTGCCGGCACATTCGCGTGTGGTCAGAGGAT 1620
 Db 521 GlyLeuAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp 540
 QY 1621 GTCACGCTGTCATGAATTAACCTCGACCGGAGTGAACACCTGATGCTGATACAA 1680
 Db 541 ValThrLeuSerLeuAsnTyrThrTrpThrArgSerGluGlnArgAspGlyAspAsnLys 560
 QY 1681 GGTGCGCGGTGATGTTATACCCCTGAACACATGTTGAATTCGAAACTGAGCAGATC 1740
 Db 561 GlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle 580
 QY 1741 ACCGAGAGGTGCATCATGCTGGTGGTGGCTTATTCGGGGAAACACACCGTTTCAC 1800
 Db 581 ThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgGlyLysThrProArgPheThr 600
 QY 1801 CAGAATTTATCTGCTACTGACGCTGTACAGAGAAGAGTGTATGATGAGAAAGGAGATAC 1860
 Db 601 GlnAsnTyrSerSerLeuSerAlaValGlnLysLysValTyrAspGluLysGlyLysThr 620
 QY 1861 CTGAAGCTTGGACGCTGTGATGAGTGTGTGTGTGGAAGATGACGATGCTCCCTGACG 1920
 Db 621 LeuLysAlaTrpThrValValAspAlaGlyLeuSerTrpLysMetThrAspAlaLeuThr 640

QY 1921 CTGAATGCTGGGTGAATTAACCTGCTCAACAGGATTACAGTACGCTGAGCTGTACAGT 1980
 Db 641 LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyrSerAspValSerLeuTyrSer 660
 QY 1981 GCCGTAAAGTACGCTGTATCCGGTGATTACTCCAGACGGGATCATCAACACAGGA 2040
 Db 661 AlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThrThrGly 680
 QY 2041 TATGTGATACCTGACGCGAAATTAATGATGTCGCTGAACTATCAGTTTC 2088
 Db 681 TyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe 696
 RESULT 3
 Q93SE1
 ID Q93SE1 PRELIMINARY; PRT; 695 AA.
 AC Q93SE1;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=4797/97;
 RA Zhang W.L.; Bielaszewska M.; Schmidt H.; Karch H.;
 RT "Characterisation of a novel pathogenicity island present in LEE-
 negative Shiga toxin-producing Escherichia coli."
 RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
 DR EMBL; AJ278144; CAC39288.1;
 DR GO; GO:0001987; C:outer membrane; IEA.
 DR GO; GO:0004872; F:receptor activity; IEA.
 DR GO; GO:0005215; P:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000531; TonB boxC.
 DR Pfam; PF00593; TonB dep_Rc; 1.
 KW Hypothetical protein; Membrane; Outer membrane; Receptor; TonB box.
 SQ SEQUENCE 695 AA; 76431 MW; 151B7053262CCEAI CRC64;

Alignment Scores:

Pred. No.: 7.6e-215 Length: 695
 Score: 3500.50 Matches: 664
 Percent Similarity: 98.42% Conservative: 21
 Best Local Similarity: 95.40% Mismatches: 10
 Query Match: 93.80% Indels: 1
 DB: 2 Gaps: 1

US-10-625-972-4 (1-2091) x Q93SE1 (1-695)

QY 1 ATCGGAATACCACTCTGGCTTCGGTAGTCATTCCTGCTCGGATTTTCAGCCAGCAGC 60
 Db 1 MetArgIleThrThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSerSer 20
 QY 61 ATACTCTCTCAGAGGATGTGATGATTCCTCGGCATCCGGCTATGAGAAAAAGCTGACT 120
 Db 21 Met---AlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLysLeuThr 39
 QY 121 AACGACCCCGCCAGTGTTCCTGTGATGTAGCCAGAGGAATTCAGTCCAGCAGTACCAC 180
 Db 40 AsnAlaAlaAlaSerValSerValIleSerGlnGluGluLeuGlnSerGlnTyrHis 59
 QY 181 GATCTGCGCGAGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGGTACGGGTAA 240
 Db 60 AspLeuAlaGluAlaLeuArgAlaValGluGlyValAspValGluSerGlyThrGlyLys 79
 QY 241 ACCGAGCGGCTGAAATCAGCATCCGAGGNAATCCGACGAGTTACAGCTGATGATGATT 300
 Db 80 ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyrThrLeuIleLeuIle 99
 QY 301 GATGGTGTTCGTCAGGCGGAGACAGTACGTCGCTGACTCCCAACGGTTTTTCTGCCATGAAT 360

Db 100 AspGlyValAlaGlnGlyGlySerGlyAspValThrProAsnGlyPheSerAlaMetAsn 119
Qy 361 ACCGGTTCATGCCCTCTCCGCTCAGCGGATTTAGGTTATCAGGGGCCGATG 420
Db 120 ThrGlyPheMetProProLeuAlaAlaAlaGluAlaGluValIleArgGlyProMet 139
Qy 421 TCCACATGTATGGCTCTGATCGCATGGCGGTGTGGTGAATATCATACAGAAAGAT 480
Db 140 SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAsnIleIleThrLysLysAsn 159
Qy 481 GCAGACAAATGGCTCTCTCCGCTCAATCAGCGGCTGATCTGCAGGAAGCAACAATGG 540
Db 160 AlaAspLysTyrLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp 179
Qy 541 GGTAAACAGCAGCCAGCTTTAAATTTCTGGAGCAGTGTCCCTTTGTGGATGATCTCTCAGC 600
Db 180 GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspAspSerValSer 199
Qy 601 CTGACGATCGCGGTAGACACACACAGCGTCAAGGTTTCATCGGTCACATCACTCAGCGAT 660
Db 200 LeuGlnIleArgGlySerThrGlnGlnArgGlnGlySerSerValThrSerLeuSerAsp 219
Qy 661 ACAGCAGGACCGCATTCTCTATCCACGAGTCACAGAAATTAATCTTTGGTGCACGT 720
Db 220 ThrSerAlaThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg 239
Qy 721 CTTGACTGGAAGCGTCGAGCAGGATGTCTCTGTTGTATGGATACACACCGCGAG 780
Db 240 IleAspTrpLysAlaThrGluGlnAspValLeuTrpPheAspMetAspThrArgGln 259
Qy 781 CATTATGATACCGGATGGCAACTGGGAGTCTGACGGGGGATATGACCGACCGCTG 840
Db 260 ArgTyrAspAsnGlnAspGlyGlnLeuGlySerLeuThrGlyGlyTyrAspHisThrLeu 279
Qy 841 CCCTATGACGGAACAAATTTTCAGCTGGCTATGATCATCTTCCACCTTCGGAACATGG 900
Db 280 ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspHisThrPheGlyThrTrp 299
Qy 901 AAATCGTATCTGAACCTGGACAGACAGAAATAAGTCTGAGCTGTGACGAGTGA 960
Db 300 LysSerTyrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal 319
Qy 961 CTGAACGCGCACAAATGGGGGTTCGCCGTGACGCGGGGAGCTTAAGGAATCGAACCTT 1020
Db 320 LeuLysSerAspLysTrpGlyLeuAlaGlyLysProArgGluLeuLysGluSerAsnLeu 339
Qy 1021 ATCTGTAATCTAATCTGTTACCCCTCTGGAGAAATCTCATCTGTTACGGTGGGGGC 1080
Db 340 IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuLeuThrValGlyGly 359
Qy 1081 GAGTTTCAGAGCTCGTCCATGAAGACGAGTGTCTTGGCAGCACAGTGAACCTTTC 1140
Db 360 GluTyrGlnSerSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe 379
Qy 1141 CGSCAGAAAAGCTGTGCGTATTGTGAGGATGAGTGCATCTCAACGGATCCATTGGG 1200
Db 380 ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspValLeuAla 399
Qy 1201 CTGACTCGGGGACCGCTATGACATCATGACANTTCGGGGACACTTCAGTCCGGT 1260
Db 400 LeuThrAlaGlySerArgTyrGluHisHisAspGlnPheGlyGlyHisPheSerProArg 419
Qy 1261 GCATATCTGCTCGGATGTGGCAGATGCTGACGCTGAAGCGGCTGTGACCAAGGA 1320
Db 420 AlaTyrLeuValTrpAspValThrAspAlaTrpThrLeuLysGlyGlyValThrThrGly 439
Qy 1321 TATAAGCCACCCAGATGGGCGAGCTACATAAGGATTAAGTGTGTCCGGCGAGGA 1380
Db 440 TyrLysAlaProThrMetGlyGlnLeuHisLysGlyIleSerGlyValSerGlyGlnGly 459
Qy 1381 AAACAAATCTACTGTGTAACCCCGACCTGAACCGGAGAGAGCTCAGTTATGAGCT 1440
:::|

Db 460 ArgThrAsnLeuLeuGlyAsnProGluLeuLysProGluGluSerValSerTyrGluAla 479
Qy 1441 GGGGTCTATTACGATAACCCCGCGTCTGAATGCCAATGTACAGTTTATGACTGAC 1500
Db 480 GlyValTyrTyrAspAsnLeuSerGlyLeuAsnAlaAsnIleThrGlyPheIleThrAsp 499
Qy 1501 TTCTCCAAACAGATCTCTCTATTCCATAAATGATAACACCAATAGCTATGTAACAGC 1560
Db 500 PheSerAsnLysIleValSerTyrSerIleAspAspAsnThrAsnSerTyrThrAsnSer 519
Qy 1561 GGAAGGCGCGGTGACCGTGTGGAATTTGCGGCGACATTCGCCGTGTGTTGTCAGAGAT 1620
Db 520 GlyLysAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp 539
Qy 1621 GTACCGTCTCATTGATACACTCGACCCGAGTGAACAGTGAACAGTGTGATGTAACAAA 1680
Db 540 ValThrLeuSerLeuAsnTyrThrTrpThrGlnSerGluGlnArgAspGlyAspAsnLys 559
Qy 1681 GGTGCGCGCTGAGTTATACCCCTGAACACATCGTGAATCGGAACCTGAACCTGGCAGATC 1740
Db 560 GlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle 579
Qy 1741 ACCGAGAGTGGCATCATGGCTGGTGGTCCCGTTATCGGGGAAAACACCCAGTTTCACC 1800
Db 580 ThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgGlyLysThrProArgPheThr 599
Qy 1801 CAGAAATTATTCGTCACCTGACGCTGTACAGAAAGTGTATGATGAGAAAGGAGAAATAC 1860
Db 600 ArgAsnTyrSerSerLeuSerAlaValGlnLysValTyrAspGluLysGlyGluTyr 619
Qy 1861 CTGAAAGCTGGACGGTGTGGATGAGGTCCTGCTGCGGAAGATGACGATGCGCTGACG 1920
Db 620 LeuLysAlaThrThrValValAspAlaGlyLeuSerTrpLysValThrAspAlaLeuThr 639
Qy 1921 CTGATCTCGGTGATTAACCTGCTCAACAAGGATACAGTCACTGACGTCGTCGTACAGT 1980
Db 640 LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyrSerAspValSerLeuTyrSer 659
Qy 1981 GCCGTAAGAGTACGCTGTATGCGGTGATTACTTCAGACCGGATCATCAACAACAGGA 2040
Db 660 AlaAspLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThrThrGly 679
Qy 2041 TATGTATCTGACGCGCAATTTACTGATGCTCGCTGACTACTCAGTTTC 2088
Db 680 TyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe 695
RESULT 4
087518 PRELIMINARY; PRT; 669 AA.
AC 087518;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Exogenous ferric siderophore receptor R4.
GN R4.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CF7073;
RX MEDLINE=9734757; PubMed=9199454;
RA Kao J.S., Stucker D.M., Warren J.W., Mobley H.L.;
RT "Pathogenicity island sequences of pyelonephritogenic Escherichia coli
CF7073 are associated with virulent uropathogenic strains."
RL Infect. Immun. 65:2812-2820(1997).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=CF7073;
RX MEDLINE=98380392; PubMed=9712795;
RA Guyer D.M., Kao J.S., Mobley H.L.T.;
RT "Genomic analysis of a pathogenicity island in uropathogenic
RT

RT Escherichia coli CFT073: distribution of homologous sequences among
 RT isolates from patients with pyelonephritis, cystitis, and catheter
 RT associated bacteriuria and from fecal samples.";
 RL Infect. Immun. 66:4411-4417(1998).
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
 DR EMBL; AF081285; AAC61730.1; -.
 DR HSRF; P03825; IEPF.
 DR GO; GO:0019857; C:outer membrane; IEA.
 DR GO; GO:0004872; F:receptor activity; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro: IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; i.
 KW Membrane; Outer membrane; Receptor; TonB box.
 SQ SEQUENCE 669 AA; 73806 MW; B559751F05B57F46 CRC64;

Alignment Scores:
 Pred. No.: 8,12e-215 Length: 669
 Score: 3500.00 Matches: 667
 Percent Similarity: 99.85% Conservative: 1
 Best Local Similarity: 99.70% Mismatches: 1
 Query Match: 93.78% Indels: 0
 DB: 2 Gaps: 0

US-10-625-972-4 (1-2091) x 087518 (1-669)

QY	82	ATGATTCTCGGATCCGGCTATGAGAAAGCTGACTAACCGACCCCGCAGTGTCTTCT	141
DB	1	MetileValSerAlaSerGlyTyrGluLysLysLeuThrAsnAlaAlaSerValSer	20
QY	142	GTGATTAGCCAGGAGAAATGCGAGTCCAGCAGTACACGATCTGCGGAGGCTCTGAGA	201
DB	21	ValileSerGlnGluGluLeuGlnSerSerGlnTyrHisAspLeuAlaGluAlaLeuArg	40
QY	202	TCAGTAGAGGCTGGATGTTGAAAGTGTACGGGTAAACCGGAGGCTGGAATCAGC	261
DB	41	SerValGluGlyValAspValGluSerGlyThrGlyLysThrGlyGlyLeuGluLeuSer	60
QY	262	ATCCGAGGATGCCAGCCAGTACACGCTGATACGATGATGATGATGATGATGATGATG	321
DB	61	IleArgGlyMetProAlaSerTyrThrLeuIleLeuIleAspGlyValArgGlnGlyGly	80
QY	322	AGCAGTACGCTGACTCCCAACGGTTTTCTGCCATGAATACCGGGTTATCGCCCTCTG	381
DB	81	SerSerAspValThrProAsnGlyPheSerAlaMetAsnThrGlyPheMetProLeu	100
QY	382	GCCGCCATTGAGGCTATGAGTTATCAGGGCCGATGTCACACTGTATGCTCTGAT	441
DB	101	AlaAlaIleGluArgIleGluValIleLeuArgGlyProMetSerThrLeuTyrGlySer	120
QY	442	GCGATGGCGGCTGTGTGATATCATATACAGAAAGATGCAGACAAATGGCTCTCTCC	501
DB	121	AlaMetGlyValValAlaIleIleThrArgLysAsnAlaAspLysTrpLeuSerSer	140
QY	502	GTCAATGCGAGGCTGAATCTCGAGGAAGCAAAATGGGGTAAACAGCAGCGATTAAT	561
DB	141	ValAsnAlaGlyLeuAsnLeuGlnGlnSerAsnLysTrpGlyAsnSerSerGlnPheAsn	160
QY	562	TTCTGCGAGCAGTGGCTCCCTGTTGGATGATTTCTGTCAGCCTCAGGTCGCGGTAGCACA	621
DB	161	PheTrpSerSerGlyProLeuValAspAspSerValSerLeuGlnValArgGlySerThr	180
QY	622	CAACAGCTGAGGTTATCGGTACATCACTGAGCGATACAGAGCGACCGGTATTCCT	681
DB	181	GlnGlnArgGlnGlySerSerValThrSerLeuSerAspThrAlaAlaThrArgIlePro	200
QY	682	TATCCACAGGAGTACAGATTAATATCTTGGTCAGCTGCTGACTGGAAGCGTCGGAG	741
DB	201	TyrProThrGlnSerGlnAsnTyrAsnLeuGlyAlaArgLeuAspTrpLysAlaSerGlu	220
QY	742	CAGGATGTCTCTGTTGATATGATACCCCGGACGCTTATGATAACCGGATGGG	801
DB	221	GlnAspValLeuTrpPheAspMetAspThrThrArgGlnArgTyrAspAsnArgAspGly	240

QY	802	CAACTGGGAGTCTCAACGGGGGATATGACCGGACCCCTGGCGTATGAGCAAAACAAAT	861
DB	241	GlnLeuGlySerLeuThrGlyTyrAspArgThrLeuArgTyrGluArgAsnLysIle	260
QY	862	TCAGCTGGCTATGATCATCTTTACCTTCGGAACATCGAATCTGATCTGAACCTGAAC	921
DB	261	SerAlaGlyTyrAspHisThrPheThrPheGlyThrTrpLysSerTyrLeuAsnTrpAsn	280
QY	922	GAGACAGAAAATAAAGGTCGTGAGCTTGTACGCACTGTACTGAAGCGCGCAAAATGGGG	981
DB	281	GluThrGluAsnLysGlyArgGluLeuValArgSerValLeuLysArgAspLysTrpGly	300
QY	982	CTTGCCGCTCAGCCGCGGAGCTTAAGGAATCGAACCTTATCTCTGAATTCATCTGCT	1041
DB	301	LeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeuIleLeuAsnSerLeuLeuLeu	320
QY	1042	ACCCTCTGGGAGAAATCTCATCTGTTACGTTGGGGGCGAGTTTCAGAGCTCGTCCATG	1101
DB	321	ThrProLeuGlyGluSerHisLeuValThrValGlyGlyGluPheGlnSerSerMet	340
QY	1102	AAAGACGAGTGTCTCTTGGCCAGCACAGGTCAAACTTCCGGCAGAAAAGCTGTGCGTA	1161
DB	341	LysAspGlyValValLeuAlaSerThrGlyGluThrPheArgGlnLysSerTrpSerVal	360
QY	1162	TTTCTGAGGATGAGTGGCATCTCACGGATGCTTCCGCTGCTGCGGCGACCGCTAT	1221
DB	361	PheAlaGluAspGluTrpHisLeuThrAspAlaLeuAlaLeuThrAlaGlySerArgTyr	380
QY	1222	GAAATCATGAGCAATTCGGGGGACACTTCAGTCCGCGTGCATATCTGCTCGGATGTG	1281
DB	381	GluHisGlnGluGlnPheGlyHisPheSerProArgAlaTyrLeuValTrpAspVal	400
QY	1282	GCAGATCCCTGGAGCTGAAAGCGGTGTGACACCGGATATAGGACCCAGCAATGGGG	1341
DB	401	AlaAspAlaTrpThrLeuLysGlyValThrThrGlyTyrLysAlaProArgMetGly	420
QY	1342	CAGTCACTAAAGGATTAGTGGTGTCTCGGGCAGGAAAAACAAATCTACTTGGTAAC	1401
DB	421	GlnLeuHisLysGlyIleSerGlyValSerGlyGlnGlyLysThrAsnLeuLeuGlyAsn	440
QY	1402	CCGACCTGAAGCGGAGAGAGCGTCAGTTATGAGCTGGGTGTATTACGATAACCC	1461
DB	441	ProAsnLeuLysProGluGluSerValSerTyrGluAlaGlyValTyrTyrAspAsnPro	460
QY	1462	GCGGCTGTAATGCAATGTACAGGTTTTATGACTGACTTCTCCACAGATGCTCTCT	1521
DB	461	AlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAspPheSerAsnLysIleValSer	480
QY	1522	TATTCCATAAATGATAACACCAATAGCTATGTAACAGCGGAAAGCGCGGTTCACGGT	1581
DB	481	TyrSerIleAsnAspAsnThrAsnSerTyrValAsnSerGlyLysAlaArgLeuHisGly	500
QY	1582	GTGGAATTCGCGGCACATTGCGCTGTGTGTGTCAGAGATGTACCGCTGTCTGCAATTAC	1641
DB	501	ValGluPheAlaGlyThrLeuProLeuTrpSerGluAspValThrLeuSerLeuAsnTyr	520
QY	1642	ACCTGGACCCGAAAGTGAACACCGTGTGTAACAAAGAGTGGCGCTGAGTTATACC	1701
DB	521	ThrTrpThrArgSerGluGlnArgAspGlyAspAsnLysGlyAlaProLeuSerTyrThr	540
QY	1702	CTGTAACACATGTTGTAATGCGAAACCTGAAGTGGCAGATACCGAAGAGTGGCATATGG	1761
DB	541	ProGluHisMetValAsnAlaLysLeuAsnTrpGlnIleThrGluGluValAlaSerTrp	560
QY	1762	CTGGTCCCGTTTATCGCGGAAAAACACACGTTTTCCACCAAGATTTATCGTCACTGAGC	1821
DB	561	LeuGlyAlaArgTyrArgGlyLysThrProArgPheThrGlnAsnTyrSerSerLeuSer	580
QY	1822	GCTGTACAGAAAGTGTATGATGAGAAAGGAGATACTGAAACCTGGACGCTGGTGG	1881
DB	581	AlaValGlnLysLysValTyrAspGluLysGlyGluTyrLeuLysAlaTrpThrValVal	600

Db 454 ProArgAlaTyrLeuValTyrAspAlaGlnAspAsnTrpThrValLysGlyGlyValSer 473
 QY 1315 ACAGGATATAGGACCCAGAGTGGCGAGCTACATAAGAGGATAGTGTGTCCGGG 1374
 Db 474 ThrGlyTyrLysThrProThrLeuAlaGlnLeuHisAsnGlyIleSerGlyValThrGly 493
 QY 1375 CAGGAGAAACAAATCTACTTGTGTAACCCGACCTGAAGCCGGAAGAGAGCGTCAGTTAT 1434
 Db 494 GlnGlyThrIleThrThrIleGlyAsnProLysLeuGluProGluSerSerValAsnThr 513
 QY 1435 GAGCTGGGTGATATACGATACCCCGCGTCTGAATCCCAATGTACAGGTTTATG 1494
 Db 514 GluValGlyValTyrGlyAsnGluThrGlyPheGlyAlaAsnValThrLeuPheHis 533
 QY 1495 ACTGACTCTCCACAAAGATTGCTCTATTCCATAAATGATAACACCAATACCTATGTA 1554
 Db 534 AsnArgPheArgAsnLysIleAsnSerValSerIleAspAsnThrThrSerThrThr 553
 QY 1555 AACAGCGAAAGCCCGGTGTCACGGTGTGGAATTTCCCGGCACATTCGCGTGTGTCA 1614
 Db 554 AsnValGlyLysAlaIleThrGlnGlyIleGluValAlaSerThrIleProLeuTyrSer 573
 QY 1615 GAGGATCTCAGCTGCTCACTGAATTCACCTGACCGAGCCGAGTGAACAACGTCATGTAT 1674
 Db 574 AspAspTrpMetLeuGlyIleAsnThrThrPheThrAspSerGluGlnLysAspGlyAsn 593
 QY 1675 AACAAAGTGGCGGTGAGTTATACCCCTGAACACATGTCGTAATCGAACTGAACTGG 1734
 Db 594 AsnLysGlyAlaArgLeuThrAsnThrProLysAsnMetValAsnAlaArgLeuAsnTrp 613
 QY 1735 CAGATACCAAGAGTGGCATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1794
 Db 614 AsnIleAsnGluGlnLeuSerThrTrpLeuLysAlaGluThrArgSerLysThrAlaArg 633
 QY 1795 TTCACCCAGAAATATTCGTCACCTGAGCGCTGTACAGAGAAAGTGTATGATGAGAAGCA 1854
 Db 634 PheThrGlnAsnTyrAlaAsnLeuSerAlaAsnLysValValTyrAsnAsnLeuGly 653
 QY 1855 GAATACCTGAAGCTGGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1914
 Db 654 SerGluPheLysProPheSerValLeuAsnLeuGlyValAlaTyrLysValThrLysAsp 673
 QY 1915 CTGACGCTGAATCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1974
 Db 674 ValThrLeuAsnGlyAlaValAsnAsnLeuLysAspPheThrArgThrHisIle 693
 QY 1975 TACAGTCCCGTGAAGTAGTACGCTGTATGCGGTGATTACTTCCAGACGGGATCATCAACA 2034
 Db 694 PheAlaValGlyAsnGlyThrThrThrAlaGlyAspPheThrSerSerGlnSerThr 713
 QY 2035 ACAGGATATGTACTGACCTGAGCAATTAATGATGTCGTCGTAATCATCAGTTTC 2088
 Db 714 AlaGlyTyrValValProGlyArgAsnTyrTrpValSerValAsnValAsnPhe 731

RESULT 6

Q8VR68 PRELIMINARY; PRT; 390 AA.
 AC Q8VR68
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE R4-like protein.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CFT073;
 RX MEDLINE=21458820; PubMed=11574920;
 RA Rasko D.A., Phillips J.A., Li X., Mobley H.L.T.;
 RT "Identification of DNA sequences from a second pathogenicity island of
 RT uropathogenic Escherichia coli CFT073: a probes specific for

RT uropathogenic populations";
 RL J. Infect. Dis. 184:1041-1049 (2001).
 CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
 DR EMBL; AF447814; AAL67380.1; -.
 DR GO; GO:0019867; C:outer membrane; IEA.
 DR GO; GO:0004872; F:receptor activity; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 KW Membrane; Outer membrane; Receptor; TonB box.
 SQ SEQUENCE 390 AA; 42718 MW; 0B72B0B02B65BA12 CRC64;
 Alignment Scores:
 Pred. No.: 6,16e-104 Length: 390
 Score: 1759.00 Matches: 343
 Percent Similarity: 99.71% Conservative: 0
 Best Local Similarity: 99.71% Mismatches: 1
 Query Match: 47.13% Indels: 0
 DB: 2 Gaps: 0
 US-10-625-972-4 (1-2091) x Q8VR68 (1-390)
 QY 1 ATGGCAATACCACTCTGGCTTCCTGATGATCCCTGCTCGGATTTTCAGCCAGCAGC 60
 Db 1 MetArgIleThrThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSerSer 20
 QY 61 ATAGCTGTCTGAGAGGATGTGATGATGTTCTCGGCATCCGCTATGAGAAAAGCTGACT 120
 Db 21 IleAlaAlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLysLeuThr 40
 QY 121 AACGACGCGCTGTTCTTCTGATAGCCAGAGGAATTCAGTCCACCCAGTACCAC 180
 Db 41 AsnAlaAlaAlaSerValSerValIleSerGlnGluLeuGlnSerSerGlnTyrHis 60
 QY 181 GATCTGCGGAGGCTCTGAGATCAGTAGAGGCTGTGATGTTGAAAGTGTACGGGTAA 240
 Db 61 AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys 80
 QY 241 ACCGAGGCTGGAATACGATCCGAGGAATGCCAGCTTACAGCTGATCATGATT 300
 Db 81 ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyrThrLeuIleLeu 100
 QY 301 GATGCTGTCTGAGCGGAGGACGATGACGTCACCTCCCAACGGTTTTTCTGGCATGAT 360
 Db 101 AspGlyValArgGlnGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
 QY 361 ACCGGGTTTCATGCCCTCTGCGCCCATTTGAGCGTATTGAGGTTATCAGGGGCGCATG 420
 Db 121 ThrGlyPheMetProProLeuAlaAlaIleGluArgIleGluValIleArgGlyProMet 140
 QY 421 TCACACTGTATGGCTCTGATGAGTGGCGGTGTGTGATATCATATTACCAAGAAAGAT 480
 Db 141 SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAsnIleIleThrArgLysAsn 160
 QY 481 GCACAAATGGCTCTCTCCGTCATGACGAGGCTGAATCTGCAGGAACCAACAATGG 540
 Db 161 AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp 180
 QY 541 GGTAAACAGCAGCAGTTTAAATTTCTGGAGCAGTGGTGGTGGTGGTGGTGGTGGTGGT 600
 Db 181 GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspSerValSer 200
 QY 601 CTGCAGTACGCGGTAGCACACACGCGTCATCGGTTCATCGGTTCATCATCATCGCGCAT 660
 Db 201 LeuGlnValArgGlySerThrGlnGlnArgGlnGlySerSerValThrSerSerAsp 220
 QY 661 ACAGCAGCAGCGGTATTCCTTATCCACGAGTTCACAGAAATTAATCTTGTGTGACGT 720
 Db 221 ThrAlaAlaThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg 240
 QY 721 CTGACTGGAGGCTCGGAGGAGTGTCTCTGTTGATATGATGATACCCCGGCGAG 780

241 LeuAspTyrLysAlaSerGluGlnAspValLeuTyrPheAspMetAspThrThrArgGln 260
 781 CTTTATGATACCGGATGGCACTGGGGAGTCTGACGGGGGATATGACCGACCGCTG 840
 261 ArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGlyTyrAspArgThrLeu 280
 841 CCGTATGAGCGAAACAAATTTTCACTGGCTATGATCATCTTTTCACTTCCGAAACATGG 900
 281 ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspHisThrPheThrPheGlyThrTyr 300
 901 AAATCGTATCTGAATCGAAGACAGACAGAAATAAAGTCGTGCTGTGACGAGTGTGTA 960
 301 LysSerTyrLeuAsnTyrAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal 320
 961 CTGAGCCGCAATGGGGCTTCGGCTGACGGCGCGGAGCTTAAGAAATCGAACCTT 1020
 321 LeuLysArgAspLysTyrGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu 340
 1021 ATCCTGAATTCA 1032
 341 IleLeuAsnSer 344
 |||||
 RESULT 7
 P70847
 ID P70847 PRELIMINARY; PRT; 754 AA.
 AC P70847;
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Exogenous ferric siderophore receptor.
 GN BFA OR B4761.
 OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alkaligenaceae; Bordetella.
 OX NCBI_TaxID=518;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271;
 RA Parthali J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
 Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
 Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 Leather S., Moulé S., Norbertczak H., O'Neil S., Ormond D., Price C.,
 Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RA "An iron-regulated outer-membrane protein specific to Bordetella
 bronchiseptica and homologous to ferric siderophore receptors."
 RL Microbiology 143:135-145(1997).
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-588;
 RX MEDLINE=22827954; PubMed=12910271;
 RA Parthali J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
 Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
 Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
 Leather S., Moulé S., Norbertczak H., O'Neil S., Ormond D., Price C.,
 Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
 RA "Comparative analysis of the genome sequences of Bordetella pertussis,
 Bordetella parapertussis and Bordetella bronchiseptica."
 RL Nat. Genet. 35:32-40(2003).
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
 DR EMBL; U56084; AAB51774.1; -;
 DR EMBL; BX640451; CAE35124.1; -;
 DR GO; GO:0019867; C:outer membrane; IEA.
 DR GO; GO:0004872; F:receptor activity; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR001064; Crystallin.
 DR InterPro; IPR000531; TonB box.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00225; CRYSTALLIN BETAGAMMA; 1.
 DR PROSITE; PS00430; TONB DEPENDENT REC 1;
 KW Membrane; Outer membrane; Receptor; TonB box; Complete proteome.

SQ SEQUENCE 754 AA; 82628 MW; A2929E37EE58AE61 CRC64;
 Alignment Scores:
 Pred. No.: 9,39e-68 Length: 754
 Score: 1192.00 Matches: 275
 Percent Similarity: 52.41% Conservative: 127
 Best Local Similarity: 35.85% Mismatches: 243
 Query Match: 31.94% Indels: 122
 DB: 19 Gaps:
 US-10-625-972-4 (1-2091) x P70847 (1-754)
 Qy 22 TCCGTAGTCAATCCCTGCTCCGATTTTCAGCCAGCAGCATAGCTGCTGCAGAG----- 75
 Db 32 AlaLeuAlaIleAlaLeuGlyAlaAlaGlnThrAlaSerProGlnGluAsp 51
 Qy 76 -----GATGTGATGATGCTCGGCATCCCGCTATGAGAAAG 114
 Db 52 ThrLeuGlyIleThrGlnMetAspThrValValThrAlaSerGlyPheGluGlnGlu 71
 Qy 115 CTGACTAACCGCAGCCGCTGTTTCTGTGATTTAGCCAGGAGAAATGCGAGTCCAGCCAG 174
 Db 72 IleLysAsnAlaProLysIleSerValIleThrArgGluGlnLeuGluSerLysPro 91
 Qy 175 TACACGATCTGGCGAGGCTCTGATCAGTACGTAGAGGTGTGATGTTGAACTGGTACG 234
 Db 92 PheHisAsnLeuAlaAspAlaValAlaAspValGluGlyValSerValGluArgGly--- 110
 Qy 235 GGTAAACCGGAGGCTGGAAATCAGCATCCGAGGAATGCCAGCAGTTACAGCTGATA 294
 Db 111 GlyLysAlaGlyGlyMetAsnIleSerIleArgGlyLeuProSerAspTyrThrLeuVal 130
 Qy 295 CTGATTCATGCTGTGCTGTGCTGATGCGATGCGGCGGTGTGGTGAATATCATTTACC 471
 Db 131 LeuValAspGlyLysArgLeuSerGlnAsnSerSerGlyAlaArgProAsnGlyPheGly 150
 Qy 352 GCCATGAATACCGGTTTCATGCCCCCTCTCGCCGCAATTCAGCGTATTGAGGTATCAGG 411
 Db 151 AspValAspThrAsnPheIleProMetSerAlaIleAspArgIleGluValValArg 170
 Qy 412 GGGCCGATGTCACACATGATGCTGTGATGCGATGCGGCGGTGTGGTGAATATCATTTACC 471
 Db 171 GlyProMetSerThrLeuTyrGlySerAspAlaMetGlyGlyValIleAsnIleIleThr 190
 Qy 472 AGAAGAATGCAGACAAATGGCTCTCTTCGTCATGCGAGGCTGATCTGCAGGAAGC 531
 Db 191 ArgLysValAlaArgGluTyrThrGlyGlnValThrLeuAspGlyThrAlaGlnGlyAsp 210
 Qy 532 AACAAATGGGTACACAGCAGCCAGTTTAATTTCTGGAGCAGTGGTCCCTTTGGATGAT 591
 Db 211 AsnArgTyrGlyAsnAsnTyrGlySerSerPheTyrLeuSerGlyProLeuGlnThrAsp 230
 Qy 592 TCTGTACGCTGACAGGTACCGGTAGCAGCACAACAGCGT-----CAGGGTTCATCG 642
 Db 231 LysLeuGlyLeuSerLeuArgGlyGlyLeuTyrArgArgLeuSerAlaHisGlySer--- 249
 Qy 643 GTCACATCATGCGCATACAGCAGGACGCGTATTCCTTATCCCGAGGAGTACAGAAAT 702
 Db 250 -----TyrProAlaAsnGlnAlaGlu 256
 Qy 703 TAT----- 705
 Db 257 TyrAspSerGlyAspTyrSerGlyAspIleAlaSerPheSerGlyLeuGlyAspSerLeu 276
 Qy 706 -----RATCTGGTGGCAGCTCTTTCATGGAAGCGCTCGGAGCAGGAGTGTCTCGTTT 759
 Db 277 GlnArgAsnValGlyLeuArgLeuAlaLeuThrProAsnArgAsnHisAspIleLeuPhe 296
 Qy 760 CATATGATACCAACCCCGCAGCTTATGATAACCGGATGGCAACTGGGAGTCTCAACG 819
 Db 297 AspValAspAlaAsnTyrGlnThrPheAspAsnAlaAsnGlyGluLeuGlyThrLeuAsn 316
 Qy 820 -----GGGGGATATGACCGGACCCCTCGCGCTATGAGCGA 852

Db 317 AlaAspValAlaProAsnArgGlnGlyGlyTyTrpGluProGluMetLysPheAsnArg 336
 QY 853 AACAAATTCA-----GCTGGCTATGATCATACTTTCACCTTCGGAACATGG 300
 Db 337 GlnArgTyrAlaLeuThrHisLeuGlyArgTyrAspGlyAlaIleSerSerAspThr--- 355
 QY 901 AATCGTATCTGAACGAGACGACGAGAAATAAAGCGTGGTCTGAGCTTGTACCGAGTGA 960
 Db 356 -----SerLeuLeuTyrAspThrThrGluThrIleGlyArgThrAsnProMetSerThr 373
 QY 961 CTGAAGCGGACAAATGGGGCTTGCCTGAGCGGAGCTTAAGAAATCGAACCTT 1020
 Db 374 ProArgGlnProSer-----AspGlyGluLysArgGluLeuGluTyrGluAsnTrp 390
 QY 1021 ATCTGAAATTCATCTGCTTACCCCTCTG---GGAGAATCTCATCTGTTAGCGTGGG 1077
 Db 391 ValPheAspThrLysTrpThrMetProLeuPheAsnAspArgHisAsnLeuThrMetGly 410
 QY 1078 GCGGAGTTTCAGAGCTCGTCCATGAAGACGAGTTGTCTTCCAGCACAGGTGAACCT 1137
 Db 411 GlyGlnTrpArgGluGlnLysPheLysAsp-----ThrLeuValSerAlaProLeuAsn 428
 QY 1138 TTCGGCAGAAAGCTGGTGGTATTGTGAGGATGAGTGGCATCTCCAGATGCACTT 1197
 Db 429 LeuArgGlnTyrGlnTrpAlaLeuPheAlaGluAspGluTrpArgIleValAspAspLeu 448
 QY 1198 GCGTGAAGTGGGCGGAGCTATGACATCATGAGCAATTCGGGGGACACATTCAGTCGG 1257
 Db 449 AlaLeuThrMetGlyAlaArgTyrAspArgAsnGluGlnPheGlyLysTrpSerPro 468
 QY 1258 CGTGCATATCTGCTGGGATGGCAGATGCTCGGACGCTGAAGCGCGGTGTGACACG 1317
 Db 469 ArgGlyTyrLeuValTrpAsnAlaThrProAlaTrpThrValLysGlyValSerLys 488
 QY 1318 GGATATAGCCACCCAGATGGGCGAGCTACATAAAGGATAGTGGTGTGCGGGCAG 1377
 Db 489 GlyTyrLysThrProAspIleAsnLeuMetThrAspGlyIleLeuGlyLeuGlyAlaGln 508
 QY 1378 GGAAGAACAAATCTACTTGTGAACCCGACCTGAAGCCGAGAGAGCGCTCAGTTATGAG 1437
 Db 509 GlyThrMetProLeuLeuGlyAsnSerGlnLeuLysProGluSerSerThrSerGlu 528
 QY 1438 GCTGGGGTGTATTACGATACCCCGCGCTGCAATGCCAATGTCACAGTTTATGACT 1497
 Db 529 LeuGlyValLeuPheAspAspGlyGlyGluGlyLeuThrGlyAsnLeuThrGlyPheHisThr 548
 QY 1498 GACTTCTCCAAACAAGATTGCTCTTATTCCTAAATGAT----- 1536
 Db 549 LysPheLysAspLysIleAspThrGlnAsnValProAsnCysLeuAlaAlaGlyGlyPro 568
 QY 1537 -----AACACC 1542
 Db 569 ValProGlyCysLeuAspLeuGlyValTrpGluArgAsnGlyValProValAlaAsnPhe 588
 QY 1543 AATAGCTATGTAACAGCGGAAAGGCCGCTGCGAGTGTGAATTTGCCGGCACATTG 1602
 Db 589 SerGlnArgValAsnValAspThrAlaThrIleGlnGlyPheGluLeuGlyGlyArgIle 608
 QY 1603 CGCTGTGTGAGAGGATGTCAGCTGTCTCACTGAATTCACCTGGACCGGAGTGAACAA 1662
 Db 609 ProLeuPhe---GluGlyTrpSerPheSerGlyAsnTyrThrLeuThrAlaSerGluIle 627
 QY 1663 CGTGATGTTGATACAAAGGTCGCGCTGAGTTATACCCCTGAACACATGTTGATCGG 1722
 Db 628 ThrSerGlyAlaLysGlnGlyGlnProLeuGlySerGlnProArgHisSerLeuAsnLeu 647
 QY 1723 AAACGAACTGGCAGATCAACGAGAGGTGGCATCATGCTGGTGGTGGCTTATCGCGGG 1782
 Db 648 GlyLeuAsnTrpArgValAlaGlnGluArgPheAsnAlaTrpValArgGlyGluTyrArgAla 667
 QY 1783 AAACACACAGTTTACCCAGAAATTTTCGTCACTGAGCGGTGTACAGAAAGATGTAT 1842

Db 668 Lys-----GlnPheAsnAspMetAsnTrpGluLysGluGlnValPhe 681
 QY 1843 GATGAGAAAGGAGAAATCCTGAAAGCCTGGACGGTGGTGGATGAGTCTGCTCGTGAAG 1902
 Db 682 -----TyrSerProTyrTrp---LeuAlaSerLeuGlyGlySerTyrVal 695
 QY 1903 ATGACGGATGCGCTGACGCTGAATGCTGCGTGAATACCTGCTCAACAAGGATTACAGT 1962
 Db 696 LeuAsnLysAsnValThrLeuSerAlaSerValTyrAsnLeuPheAspLysAsnPheVal 715
 QY 1963 GACCTGAGCCCTGTACAGTGGCGGTAAAGAGT-----ACGCTGTATGCGCGT 2007
 Db 716 AspTyrGlyProThrLysValGlyThrSerAlaProThrAlaAlaThrSerTyrSerAsn 735
 QY 2008 GATTACTTCCAGAGCGGATCATCAACACAGGATATGTTGATACCTGAGCGAAATTACTGG 2067
 Db 736 SerTyrArgGln-----ValLeuGluGlyArgArgLeuTrp 747
 QY 2068 ATGTCGCTGAACTATCATGTTTC 2088
 Db 748 ValSerAlaAsnIleThrPhe 754
 RESULT 8
 Q9PPG3 PRELIMINARY; PRT; 696 AA.
 AC Q9PPG3;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative iron uptake protein.
 GN CFRA OR C00755.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Campylobacteraceae; Campylobacter.
 OX NCBI_TaxID=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCCT 11168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
 RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrall B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
 RT reveals hypervariable sequences."
 RL Nature 403:665-668(2000).
 DR EMBL; AL139076; CAB73021.1; -
 DR PIR; D81346; D81346.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004872; F:receptor activity; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000531; TonB_boxC.
 KW Pfam; PF00593; TonB_dep_Rac; 1.
 KW Complete proteome.
 SQ SEQUENCE 696 AA; 77554 MW; 31E09CB15187388C CRC64;
 Alignment Scores:
 Pred. No.: 1,79e-62 Length: 696
 Score: 1109.00 Matches: 255
 Percent Similarity: 54.80% Conservative: 150
 Best Local Similarity: 34.51% Mismatches: 248
 Query Match: 29.72% Indels: 86
 DB: 16 Gaps: 16
 US-10-625-972-4 (1-2091) x Q9PPG3 (1-696)
 QY 1 ATCGGATAACCACTCTGGCTCCGTAGTATCCCTGCTCGGATTTTCGCCAGCAGC 60
 Db 1 MetLysLysIleCysLeuSerValCysAlaIleGlyLeuLeuSerSerAsnAlaIleSer 20
 QY 61 ATAGCTGCTCAGAGGATGTGATGTCTCGGCATCCGGCTATGAGAAAAGTGACT 120

Qy	1021	ATCTGTAATCAITACTGCTTACCCCTCTGGGAGAATCTCATCTGGTTACGGTGGGGGGC	1080
Db	350	ileLeuAspThrLysSerValIleProLeuGlyGlnSerHisLeuSerValArgGly	369
Qy	1081	GAGTTTCAGAGCTCGTCATCAAGACGGAGTTCCTCTCCACGACAGGTGAACCTTTC	1140
Db	370	GluTyrArgLeuGluLysMetGlnAspLysIle-----AlaSerProThrAsnPhe	386
Qy	1141	CGGACAGAAAGCTGCTCGTATTCTGTGAGATGAGTGGCATCTCACGGATGCACCTTGC	1200
Db	387	AspGlnTyrLeuLeuAlaIlePheAlaGluAspGluTyrSerIleLysAspLeuArg	406
Qy	1201	CTGACTCGCGGACCGCTTAACATCATCAGCAATTCCGGGGACACTTCAGTCGCGCT	1260
Db	407	LeuThrPheGlyAlaArgTyrAsnHisIleGluIlePheGlyAsnAsnValSerProArg	426
Qy	1261	GCATATCTGCTGGGATGTGGCAGATGCTCGACGCTGAAAGCGCGTGTGACACGGGA	1320
Db	427	AlaTyrValValTyrAsnProThrAsnGluLeuThrLeuLysGlyGlyValSerThrGly	446
Qy	1321	TATAAGGCCACCCAGAAATGGGCGAGCTACATAAAGGGATTAGTGGTGTCCCGGACGGA	1380
Db	447	PheArgThrProTyrAlaAsnArgLeuIleAsnGlyThrTyrSerTyrSerGlyGlnGly	466
Qy	1381	AAACAAATCTACTTGTGTAACCCCGACCTGAAGCCGGAAGAGCGGTCACTTATAGGCT	1440
Db	467	ArgPheProThrTyrGlyAsnProAspLeuLysGluGluThrSerLeuAsnTyrGluIle	486
Qy	1441	GGGTGTATTACCATAAACCCCGCGTCTGAATGCCAAATGTCACAGGTTTATGACTGAC	1500
Db	487	AlaAlaIleTyrAsnAsnAspLeu---PheTyrValSerAlaThrGlyPheLeuThrAsn	505
Qy	1501	TTCTCCAAAGATGTCTTATTCCATAAAT-----	1533
Db	506	PheLysAspLysIleSerSerGlnSerTyrAsnAsnSerGluProIleProGlyIleGly	525
Qy	1534	-----CATACACCAATCATGTATAACAGCGGAAGGCCCGGTGACCGT	1581
Db	526	ThrCysAspAlaAspArgCysSerArgAlaIleAsnHisGlyLysValGluTyrLysGly	545
Qy	1582	GTGGAATTT---CCCGGCACATTGCCGCTGTGGTCAGAGGATGTCACGTGTCACTGAAT	1638
Db	546	ValGluLeuGlyAlaGlyIleSerProLeu-----AspAsnLeuAsnValAsnPheAla	563
Qy	1639	TACACTGGACCCGAGTGAACACCTGTATGGTGATACAA-----GGTGGCCG	1689
Db	564	TyrThrTyrLeuAspThrGluValLysGluAlaGlnAspArgSerValIleGlyLysPro	583
Qy	1690	CTGAGTTATACCCCTGAACACATGGTGAATGCGAAATCGAACTGGCAGATCACCGAAG	1749
Db	584	GluGlnAspSerLeuLysHisAsnIleMetLeuLysThrGluTyrSerPheTyrAsnLys	603
Qy	1750	GTGGCATCATGGCTGGTGCC-----CGTTATCGCGGGAACACACCATGT	1794
Db	604	PheThrProTrpIleLysGlyGluTrpGlnIleAspArgTyrMetGlyAspThr-----	621
Qy	1795	TTCAACCCAGAAATTATTTCGCTACTGAGCGGTGTACAGAGAAAGTGTATCATGAGAAAG	1854
Db	622	-----AsnIleAsnArg	625
Qy	1855	GAATACCTGAACCTGGACGGTGGATGACGAGTCTCTCTGTGGNAGATGACGATGCC	1914
Db	626	GluTyrTyrLysAspIlePheLeuAlaSerMetGlyValArgTyrAspIleAsnLysGln	645
Qy	1915	CTGACGCTGAATGCTCGGTGTAATCACTGCTCAACAGGATATACAGTGCAC---GTGAGC	1971
Db	646	TrpSerIleSerAlaAlaIleTyrAsnLeuPheAspAsnSerPheThrAsnGlyTrpGlu	665
Qy	1972	CTGTACAGTGCCTGGTAAGAGTACGCTGTATGCGGTGATTACTTCCAGACGGGATCATCA	2031
Db	666	SerTyrAlaSerGlySerGlySerThrTrpValAsnThrTyr-----	679

Qy	2032	ACACAGGATGATGATACCTGACCGGAATTA	CGTATCGCTGCTACATCACTTC	2088
		+++		
Db	680	-----AsnA:gileGluGluGlyArgArgMetTyIleSerIleAsn	Phe	696
RESULT 11				
Q8ZG14				
ID	Q8ZG14	PRELIMINARY;	PRT;	665 AA.
AC	Q8ZG14;			
DT	01-MAR-2002	(TREMBLrel. 20, Created)		
DT	01-MAR-2002	(TREMBLrel. 20, Last sequence update)		
DT	01-JUN-2003	(TREMBLrel. 24, Last annotation update)		
DE	Putative outer membrane protein (Putative outer membrane			
DE	iron/siderophore receptor).			
GN	VP01313	OR YIUR OR Y2872.		
OS	Yersinia pestis.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Yersinia.			
OX	NCBI_TaxId=632;			
	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=CO-92 / Biovar Orientalis;			
EX	MEDLINE=21470413; PubMed=11586360;			
RA	Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,			
RA	Pearlman M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,			
RA	Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,			
RA	Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,			
RA	Peltwell T., Hamlin N., Holroyd S., Ogilvie K., Kariyasek A.V.,			
RA	Leather S., Moulé S., Oyston P.C.F., Quail M., Rutherford K.,			
RA	Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;			
RT	"Genome sequence of Yersinia pestis, the causative agent of plague."			
RL	Genome 413:523-527(2001).			
	[2]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=KIMS / Biovar Mediaevalis;			
EX	MEDLINE=22137863; PubMed=12142430;			
RA	Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,			
RA	Perna N.T., Rose D.J., Mau B., Zhou S., Schwarz D.C.,			
RA	Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,			
RA	Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,			
RA	Perry R.D.;			
RT	"Genome sequence of Yersinia pestis KIM.";			
RL	J. Bacteriol. 184:4601-4611(2002).			
DR	EMBL; AF414147; CAC90143.1; -			
DR	EMBL; AE013889; AAM86423.1; -			
DR	FIR; AD0160; AD0160.			
DR	GO; GO:0016020; C:membrane; IEA.			
DR	GO; GO:0004872; F:receptor activity; IEA.			
DR	GO; GO:0005215; F:transporter activity; IEA.			
DR	GO; GO:0006810; P:transport; IEA.			
DR	InterPro; IPR000531; TonB_boxC.			
DR	Pfam; PF05939; TonB_dep_Rec; 1.			
DR	PROSITE; PS00430; TonB_DEPENDENT_REC_1; 1.			
SW	Hypothetical protein; Receptor; Complete proteome.			
QK	Proteosomal protein; Receptor; Complete proteome.			
SK	SEQUENCE	665 AA; 73909 MW; F5ACA23F30B06B71 CRC64;		

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Iron-regulated outer membrane virulence protein homolog.
 GN VP2602.
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OC NCBI_TaxID=670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RMD 2210633 / Serotype O3:K6;
 RA MEDLINE=22508454; PubMed12620739;
 RX Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yaeunaga T., Honda T., Shinagawa H., Hattori M., Iida T.,
 RT "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism
 RT distinct from that of *V. cholerae*."
 RL Lancet 361:743-749 (2003).
 DR EMBL; AF005082; BAC60865.1;
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0004872; F:receptor activity; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Complete proteome.
 SQ SEQUENCE 552 AA; 71947 MW; D05C2AA617508C9E CRC64;

Alignment Scores:

Pred. No.: 7,71e-57 Length: 652
 Score: 1020.50 Matches: 255
 Percent Similarity: 52.57% Conservative: 123
 Best Local Similarity: 35.47% Mismatches: 232
 Query Match: 27.34% Indels: 109
 Gaps: 24

US-10-625-972-4 (1-2091) x Q87LK9 (1-652)

QY 16 CTGGCTTCCTAGTCAATTCCTCTCGGATTTTCAGCCAGCAGCATAGCTGCTGCAGAG 75
 DB 15 LeuAlaPheValSerValProHisVal---PheAlaAsnAspSerValSerLysValGlu 33
 QY 76 GATGTGATGATTTCTCGGCATCCGCTATGAGAAAGAAAGTCACTAAGCAGCCGCCAGT 135
 DB 34 ThrVal---ValValThrAlaSerSerTyrgluInSerGlnAlaAspAlaProAlaSer 52
 QY 136 GTTTCGTGATTAAGCAGGAGGATTCAGTCCAGCCAGTACCATGCTGCGGAGGCT 195
 DB 53 IleSerValIleSerArgGluLeuLeuAspSerArgTyrgluValThrAspAla 72
 QY 196 CTGAGATCAGTACGAGGCTGATGTTGAAAGTGTACGGTAAACCGGAGGCTGGAA 255
 DB 73 LeuLysSerValProGlyValValValThrGlyGlyGlyAspThrThr-----Asp 89
 QY 256 ATCAGATCCGAGGATCCAGCCAGTTACAGCTGATCACTAATGATGAGGTTGCTGAG 315
 DB 90 IleSerIleArgGlyMetGlySerLysTyrgluValIleLeuValAspGlyLysArgGln 109
 QY 316 GCGGGAAGCAGTACGATGCTCCCAACGGTCTTCTC---GCCATGAATACCGGTTTCATG 372
 DB 110 ---SerThrArgGluThrArgProAsnSerAspGlyProGlyIleGluGlnGlyTrpLeu 128
 QY 373 CCCCCTCTGGCCGCAATGAGGTATGAGGTATACAGGGGCCGATGTCACATGAT 432
 DB 129 ProProLeuGlnAlaIleGluArgIleGluValIleArgGlyProMetSerThrLeuTy 148
 QY 433 GGCCTCTGATGCGGCGGCTGGTGGATATCAITACCAGAAAGAAATGCAGACAAATGG 492
 DB 149 GlySerAspAlaIleGlyValIleAsnValIleThrArgLysAspAlaLeuGluTrp 168
 QY 493 CTCTCTTCGTCATGCGGCTGATCTGCGAGGAAGCAAAATGGGGTAAACAGCAGC 552

DB 169 ThrGlyAsnValGlnLeuGlyThrValIleGlnIleAsnSerArgSerGlyGlyGln 189
 QY 553 CAGTTTAAATTTCTGAGCAGGTGTCCTCTTGTGGATGATTCTGTCAACCTCGAGGTACGC 612
 DB 189 SerAlaAsnPheValAsnGlyProLeuAla---GluAsnLeuLeuGlnValTy 207
 QY 613 GGTAGCACACAACAGCGTTCATCGGTTCATCGGTTCATCGGTTCATCGGTTCATCGGTTC 672
 DB 208 GlyGlnTyThrAlaArgGluGluAspAsp----- 217
 QY 673 CQTATTCCTTCCACCGGAGTCCACAGAAT---TATAATCTTGTGTGACGCTTTGACTGG 729
 DB 218 ---IleAspTyrglyTyrgluAspLysAspMetGlnSerIleSerSerLysLeuIleTy 236
 QY 730 AAGCGCTCGGAGCAGGATGCTCTGTTTGTATGATGATACCAACCGGCGAGCTTATGAT 789
 DB 237 GlnIleAsnAspArgHisSerValGlnLeuGluGlyThrSerAlaGln----- 253
 QY 790 AACCGGATGGGCAACTG-----GGGAGTCTGACGGGGGATAT 828
 DB 254 SerArgGlyAsnValGlyLeuSerValProThrThrGlyCysArgArgGlyCysGlu 273
 QY 829 GACCGACCTCGCTATGAGCGGAAACAAATTTACGCTGGCTATGATCATATCTTCACC 888
 DB 274 AspSerLeuAsnGluTyrgluArgAsnTyrgluValThr-----LeuSerHisThr----- 289
 QY 889 TTCGGAACATCGAAA-----TCGTATCTGAACCTGGAACGACGACGACGACGAA 930
 DB 290 ---GlyGluTrpGluLeuLeuGlyAsnSerAspThrTyrgluGlnArgGluSerGlu 308
 QY 931 AATAAAGCTGCTGAGCTTGTACGCGAGTGTACTGAAGCGGACAAATGGGGCTTGCCTG 990
 DB 309 AsnLysSerArgGluMetThr----- 315
 QY 991 CAGCGCGGAGCTTAAGGAATCGAACCTTATCTGTAATCTACTTACTTACCTCTG 1050
 DB 316 -----IleValAsnThrThrPheLysSerSerLeuValLysGlyLeu 329
 QY 1051 GGAGAAATCTCATCTGCTAGCGTGGGGCGGAGTTTCAGAGCTCGTCCATGAAGACGGA 1110
 DB 330 Gly---AlaHisThrLeuThrThrGlyValAspAlaThrHisAlaGluLeuGluAspPhe 348
 QY 1111 GTTCTCCTCCGACGACAGGTGAACCTTCCGCGCAGAAAGC---TGGTCGGTATTGCT 1167
 DB 349 ThrSerAsnLysAlaSerLysThrLysAlaSerAsnThrGlnTrpAlaValPheIle 368
 QY 1168 GAGATCAGTGGCATCTCAGGATGCACTTGCCTGCTGCTGCGGCGAGCGCTATGAACAT 1227
 DB 369 GluAspGluTrpLysIleAlaGluProPheSerLeuThrLeuGlyGlyArgLeuAspHis 388
 QY 1228 CATGAGCAATTCGGGGGACACTTCAGTCCGCTGCATATCTGCTCTGGATGTGGCAGAT 1287
 DB 389 AspGluAsnTyrglyAlaHisPheSerProArgValTyrglyValTrpArgValAlaPro 408
 QY 1288 GCTTGGACGCTGAAGCGCGTGTGACCGGGATATAAGCCACCCAGATGGGGCAGCTA 1347
 DB 409 AlaTrpThrValLysGlyGlyValAlaThrGlyPheArgSerProGlnLeuArgGluIle 428
 QY 1348 CATAAAGGGAATAGTGTGCTCGCGGCGAGGAAACAAATCTACTTGTAAACCCCGAC 1407
 DB 429 ThrProGlyTrpAlaGlnValSerGlyGlyGly-----AsnIleTyrglyAsnProAsp 446
 QY 1408 CTGAGCCGGAAGAGCGCTCAGTTATGAGGTGGGTGTTATACATAAACCCCGCGT 1467
 DB 447 LeuAspProGluThrSerLeuAsnLysGluIleSerValLeuTyrgluGlnGlyAspSerGly 466
 QY 1468 CTGAAGGCCAATCTCAGAGTTTATGACTTCTCCAAACAAAGATTCTC----- 1518
 DB 467 LeuAspValThrLeuThrAlaPheHisAsnGluPheLysAspLysIleThrArgValVal 486
 QY 1519 -----TCTTATTCATAAATGATAACACCAATAGCTAT----- 1551
 DB 487 CysProAspThrIleCysThrAspGlyProAsnGlnTrpGlyAlaAspProThrTyrg 506


```

QY 1552 GTAAACACGCGAAAGCCCGGTTGACAGCGTGTGAATTTGCCGACATTCGCGCTGTGG 1611
Db 507 ILeuValAspGluAlaValThrGlnGlyValGluAlaThrLeuAlaLysProLeu--- 525
QY 1612 TCAGAGGATGTCACCGCTGCTCACTGAATATACACCTGGACCCGAGTGAACACGTTGGT 1671
Db 526 ThrGluThrIleTyrLeuSerSerTyrThrPheThrAspSerGluGlnLysThrGly 545
QY 1672 GATACAAAGGTGGCCGCTGAGTATACCCCTGAACACATGGTGAATGCCAACTGAAC 1731
Db 546 GluTyrLysGlyMetProLeuGlnLeuProLysHisLeuPheAsnValAspValThr 565
QY 1732 TGGCAGATCACGGAAGAGTGCATCATGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1791
Db 566 TrpGlnThrThrAspAsnLeuSerTrpThrLysValThrTyrArgGlyLysGluMet 585
QY 1792 CGTTTCAACCCAGAAATATTGCTACTGAGCCCTGTACAGAAAGTGTATGATGAGAAA 1851
Db 586 AspProValThrGlyProSerArgAsnSerIleValGlu----- 598
QY 1852 GGAGATACCTGAAGCCTGGACGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1911
Db 599 -----ProAlaThrThrPheValAspAlaGlyValThrTyrGlnLeuThrAsp 614
QY 1912 GCCCTGACGCTGAATGCTGCGGTGAATTAACCTGCTCAACAGGATTACAGTGCAGTGCAGC 1971
Db 615 AsnThrLysIleLysGlyAlaIleTyrAsnLeuPheAspGluAspIleAsn----- 631
QY 1972 CTGTACAGTCCGGTAAAGATGACGCTGTATGCCGTGATTAATCTCCAGCGGATCATCA 2031
Db 632 ---TyrLys-----GluTyr----- 635
QY 2032 ACAACAGATATGTATACCTGAGCGAAATTAATCTGCTGGTGAACATATCAGTTC 2088
Db 636 -----GlyTyrValGluAspGlyArgTyrTrpLeuGlyLeuAspValAlaPhe 652

```

RESULT 13

Q93RD3 ID Q93RD3 PRELIMINARY; PRT; 652 AA.

AC STRAIN=WPI;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Iron-regulated outer membrane virulence protein homolog.

GN IRGA.

OS Vibrio parahaemolyticus.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

OC Vibrionaceae; Vibrio.

OX NCBI_TaxID=670;

RN [1]

RP SEQUENCE FROM N.A.

RA Tanabe T., Funahashi T., Miyoshi S., Yamamoto S.;

RT "Identification and characterization of the Vibrio cholerae irgAB

RT homologs in Vibrio parahaemolyticus".

RL Submitted (DEC-2000) to the EMBL/GenBank/DBSJ databases.

CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).

DR EMBL; AB052655; BAB60695.1; -.

DR GO; GO:0019867; C:outer membrane; IEA.

DR GO; GO:004872; F:receptor activity; IEA.

DR GO; GO:005215; F:transporter activity; IEA.

DR GO; GO:000810; P:transport; IEA.

DR InterPro; IPR000531; TonB_box.

DR Pfam; PF00593; TonB_dep Rec; 1.

DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.

DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.

KW Membrane; Outer membrane; Receptor; TonB box.

SQ SEQUENCE 652 AA; 71957 MW; ACB945161D926FD3 CRC64;

Alignment Scores:

Pred. No.: 1.39e-56 Length: 652

Score: 1016.50 Matches: 255

Percent Similarity: 52.43% Conservative: 122

Best Local Similarity: 35.47% Mismatches: 233

Query Match: 27.24% Indels: 109

DB: 2 Gaps: 24

US-10-625-972-4 (1-2091) x Q93RD3 (1-652)

```

QY 16 CTGGCTTCGGTATGATCATTCCTCTCGGATTTTCAGCCAGCAGCATACTGCTGCAGAG 75
Db 15 LeuAlaPheValSerValProHisVal---PheAlaAsnAspSerValSerLysMetGlu 33
QY 76 GATGTGATGATGTCCTCCGATCCCGCTATGACAAAAAGCTGACTAACCCAGCCCGCAGT 135
Db 34 ThrVal---ValValThrAlaSerSerTyrGluGlnSerGlnAlaAspAlaProAlaSer 52
QY 136 GTTCTGTGATATAGCCAGGAGGAATTCAGTCCAGCCAGTACCAGCATCTGGCGGAGGCT 195
Db 53 IleSerValIleSerArgGluGluLeuAspSerArgTyrTyrArgAspValThrAspAla 72
QY 196 CTGAGATCAGTAGAGGCTGTGAATGTTGAAAGTGTACCGGTAAACCCGAGGCGCTGGAA 255
Db 73 LeuLysSerValProGlyValValValThrGlyGlyGlyAspThrThr-----Asp 89
QY 256 ATCAGATCCGAGGAATGCCAGCCACTTACACGCTCATCTGATGATGCTGTTCTCTCAG 315
Db 90 IleSerIleArgGlyMetGlySerLysTyrThrLeuIleLeuValAspGlyLysArgGln 109
QY 316 GCGGAGACGATGACGTGACTCCCAACGGTTTTCT---GCCATGAATACCCGGGTTCAATG 372
Db 110 ---SerThrArgGluThrArgProAsnSerAspGlyProGlyIleGluGlnGlyTrpLeu 128
QY 373 CCCCTCTGCCGCCCATTCAGCGTATGAGTTATCAGGGGCGCGATGCCACACTGTAT 432
Db 129 ProProLeuGlnAlaIleGluArgIleGluValIleArgGlyProMetSerThrLeuTyr 148
QY 433 GGTCTCTGATCCGATGGCGGTGTGTGAATATCATTCACGAAAGATGCAGACAAATGG 492
Db 149 GlySerAspAlaIleGlyValIleAsnValIleThrArgLysAspAlaLeuGluTrp 168
QY 493 CTCTCTTCGTCATCCAGGCTGAATCTGCAGAAAGCAACAATGGGTACACAGCAGC 552
Db 169 ThrGlyAsnValGlnLeuGlyThrValIleGlnGluAsnSerArgSerGlyGlyGluGln 188
QY 553 CAGTTTAATTTCTGGAGCAGTGTGCTCCCTGTGGATGATCTCTCAGCCCTGCGAGTACGC 612
Db 189 SerAlaAsnPhePheValAsnAlaProLeuAla---AspAsnLeuLeuGlnValTyr 207
QY 613 GGTAGCACACACAGCGCTCAGGTTTCATCGGTCACTCATCTAGCCGATACAGCAGGCAGC 672
Db 208 GlyGlnTyrThrAlaArgGluGluAspAsp----- 217
QY 673 CGTATTCCTTATCCAGGAGTCACAGAAT---TATAATCTTGGTGCACGCTCTTGACTGG 729
Db 218 ---IleAspTyrGlyTyrGluAspLysAspMetGlnSerIleSerSerLysLeuIleTyr 236
QY 730 AAGCGCTCGGAGCAGGATGCTCTGTGTTTATGATATGATACCCACCGGACGCTTATGAT 789
Db 237 GlnIleAsnAspArgHisSerValGlnLeuGluGlyGlyThrSerAlaGln----- 253
QY 790 AACCGGATGGCAACTG-----GGGAGCTGACCGGGGGATAT 828
Db 254 SerArgArgGlyAsnValGlyLeuSerValProThrThrGlyCysArgArgGlyCysGlu 273
QY 829 GACCGGACCTCGCTATGAGCGAAACAAAATTTTCAGTGGCTATGATCATCTTTTCCACC 888
Db 274 AspSerLeuAsnGluTyrArgArgAsnTyrValThr-----LeuSerHisThr----- 289
QY 889 TTCGAAACATGGAAA-----TCGTATCTGAACCTGGAACAGACAGAA 930
Db 290 ---GlyGluTrpGluLeuLeuGlyAsnSerAspThrTyrLeuGlnArgGluGluSerGlu 308
QY 931 AATAAAGGTGCTGAGCTTGACGAGTGACTGAAGCGCGACAAATGGGCGCTTGCCTGGT 990

```


Db 309 AsnLysSerArgGluMetThr----- 315
QY 991 CAGCCGGGAGCTTAAGGAATCAACCTTATCTGAATTCATTACTGTTACCCCTCTG 1050
Db 316 -----11leValAsnThrThrLeuLysSerSerLeuValLysGlyLeu 329
QY 1051 GGGAATCTCATCTGTTAGCTGGGGCGAGTTTCAGAGCTCTCCATGAAGACGA 1110
Db 330 Gly-----AlaHisThrLeuThrThrGlyValAlaThrHisAlaGluLeuGluAspPhe 348
QY 1111 GTTGTCTCTCCAGCAGCAGTGAATCTTCCGACAGAAAAGC---TGGTCGGTATTGCT 1167
Db 349 ThrSerAsnLysAlaSerSerLysThrLysAlaSerAsnThrGlnTrpAlaValPheIle 368
QY 1168 GAGATGAGTGGCATCTCAGGATGACCTGCGCTGACTCGCGGAGCGCTATGAACAT 1227
Db 369 GluAspGluTrpLysIleAlaGluProPheSerLeuThrLeuGlyLysArgLeuAspHis 388
QY 1228 CATGAGCAATTCGGGGACACTTCAGTCCGCGTGCATATCTGCTCGGATGTGGCAGAT 1287
Db 389 AspGluAsnThrGlyAlaHisPheSerProArgValThrGlyValTrpArgValAspPro 408
QY 1288 GCTGAGCGTGAAGCGGTGTGACCGGGATATAAGCCACCCAGAAATGGGCGACTA 1347
Db 409 AlaTrpThrValLysGlyGlyValAlaThrGlyPheArgSerProGlnLeuArgGluIle 428
QY 1348 CATAAAGGATTAGTGGTGTCCGGCAGGAGGAAACAAATCTACTCTGTAACCCCGAC 1407
Db 429 ThrProGlyTrpAlaGlnValSerGlyGly-----AsnIleTyroGlnGlyAspSerGly 446
QY 1408 CTGAAGCGGAGAGCGTCACTATGAGCTGGGGTGTATTACGATAACCCCGCGT 1467
Db 447 LeuAspProGluThrSerLeuAsnLysGluIleSerValLeuTyroGlnGlyAspSerGly 466
QY 1468 CTGAATGCCAATCTCAGGTTTATGACTGCTTCTCCAAACAAGTTGTC----- 1518
Db 467 LeuAspValThrLeuThrAlaPheHisAsnGluPheLysAspLysIleThrArgValVal 486
QY 1519 -----TCTATTCCATAATGATAACCAATAGTACTAT----- 1551
Db 487 CysProAspThrIleCysThrAspGlyProAsnGlnTrpGlyAlaAspProThrTyroArg 506
QY 1552 GTAAACAGCGGAAGCGCGGTGTCACGCTGGGAATTTCCGGCAGCATGTCGCTGG 1611
Db 507 IleAsnValAspGluAlaValThrGlnGlyValGluAlaThrLeuAlaLysProLeu--- 525
QY 1612 TCAGAGGATGTCACGCTGTCACTGATTTACCTGACCGGAGTGAACCAACGTCATGCT 1671
Db 526 ThrGluThrIleTyroLysSerSerSerThrThrPheThrAspSerGluGlnLysThrGly 545
QY 1672 GATAACAAGGTGGCGCTGAGTTATACCTGTAACACATCGTGAATCGGAACCTGAAC 1731
Db 546 GluTyroLysGlyMetProLeuGlnGlnLeuProLysHisLeuPheAsnValAspValThr 565
QY 1732 TGSCACATCAGCAAGAGTGCATCATGCTGCGTGGGTCGCGTATCGGGGAAACACCA 1791
Db 566 TrpGlnThrThrAspAsnLeuGluSerTrpThrLysValThrTyroArgGlyLysGluMet 585
QY 1792 CGTTTCCACCAAGATTTATTCGCTCACTGACGCGTGTACAGAGAAAAGTGTATGATGAGAA 1851
Db 586 AspProValThrGlyProSerArgAsnSerIleValGlu----- 598
QY 1852 GGGAATACCTGAAGCTGGAGCGGTGGTGGATGAGGTCTCTCTGGAAGATGACGAT 1911
Db 599 -----ProAlaTyroThrPheValAspAlaGlyValThrTyroGlnLeuThrAsp 614
QY 1912 GGCCTGACCTGAATCTCGGTGAATACCTGCTCAACAAGATTAACAGTACGACGTGAGC 1971
Db 615 AsnThrLysIleLysGlyAlaIleTyroAsnLeuPheAspGluAspIleAsn----- 631
QY 1972 CTGTACAGTGGCGGTGAAGAGTACGCTGTATGCGGTGATTACTCCAGACGGGATCATCA 2031
Db 632 ---TyrLys-----GluTy----- 635

QY 2032 ACAACAGATATGTGATACCTGACGGAATTTACTGGATGTGGTGAACATATCAGTTC 2088
Db 636 -----GlyTyroValGluAspGlyArgArgTyroTrpLeuGlyLeuAspValAlaPhe 652
RESULT 14
Q8U914 PRELIMINARY; PRT; 707 AA.
ID Q8U914;
AC Q8U914;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 25, Last annotation update)
DE Exogenous ferric siderophore receptor.
GN BFAA OR ATU3816 OR AGR_L1858.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McLelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58.";
RL Science. 294:2317-2323 (2001).
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmieu K., Gordon J., Vaudin M., Tarchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursen J., Lono C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328 (2001).
DR EMBL; AE009322; AAL44724.1; ALT_INIT.
DR EMBL; AE008292; AAK89503.1; -
DR PIR; AF3038; AF3038.
DR PIR; E98247; E98247.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001005; Myb DNA binding.
DR InterPro; IPR00531; TonB boxC.
DR Pfam; PF00593; TonB dep.Rec; 1.
DR PROSITE; PS00037; MYB_1; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 707 AA; 77944 MW; E33FC510CPA52D69 CRC64;

Alignment Scores: 4.43e-55 Length: 707
Pred. No.: 993.00 Matches: 242
Score: 49.80% Conservative: 133
Percent Similarity: 32.14% Mismatches: 244
Best Local Similarity: 26.61% Indels: 134
Query Match: 16 Gaps: 23
DB:

US-10-625-972-4 (1-2091) x Q8U914 (1-707)

QY 19 GCTTCGGTAGTCATTCCCTGCTCGGATTTTCAGCCAGCAGCATAGCT-----GCTGCA 72
Db 26 AlaAlaMetAlaGlyThrAlaLeuGlyPheAlaValProAlaPheAlaGlnAlaSer 45
QY 73 GAG-----GATGTGATGATGCTCGGCATCCGGCTATGAGAAAGCTG 117
Db 46 GluGlyAsnThrValLeuGlnGlnIleValValThrAlaSerGlyPheGluGlnAsnVal 65
QY 118 ACTAACGCGCGCGGAGGCTTCTGTGATAGCAGGAGGAATTCAGCTCAGCCAGTAC 177
Db 66 LysAspAlaProAlaSerIleThrValThrArgGluAspLeuGluLysGlySerTyr 85
QY 178 CAGCATCTGGCGGAGGCTCTGAGATCAGTACAGAGGTGTGGATGTGAAGTGTACGGGT 237
Db 86 ArgAspLeuThrAspAlaLeuArgGluValGlnGlyValSerVal-----ThrGly 102
QY 238 AAAACCGGAGGCTGGAATCAGATCCGAGGAATCCAGCCAGTTACACCTCATACTG 297
Db 103 IleAlaAsnGluLysAspValPheIleArgGlyLeuProGlyAlaTyrThrLeuIleLeu 122
QY 298 ATTGATGCTGTTCGTCAGGCGGAGGAGCAGTACGCTGACTCCCAAGCGTTTTCGCCATG 357
Db 123 ValAspGlyLysArgGln---SerThrArgAspAlaArgThrAsnGlyAsnSerGlyPhe 141
QY 358 AATACCGGGTTCATCCCGCTCTGCGCCGCTATGAGCGTATGAGCTTATCAGGGGCGG 417
Db 142 GluGlnSerPheValProValSerAlaIleGluArgIleGluValValArgGlyPro 161
QY 418 ATGTCACACATCTATGGCTCTGATCGGATGGCGGTGTGGTGAATCATTCATACAGAAAG 477
Db 162 MetSerSerLeuTyrGlySerAspAlaMetGlyGlyValIleAsnIleIleThrArgLys 181
QY 478 AATCAGACAAATGGCTCTCTCCCTCAATGACGGGCTGAATCTCCAGGAAGAACAAA 537
Db 182 ValGlyAspValTrpSerGlySerValThrThrGluGlyThrValGlnGlnHisSerLys 201
QY 538 TGGGTAACAGCAGCGCTTAAATTCGACAGTGTCCCTTGTGATGATCTCTGTC 597
Db 202 PheGlyAsnSerGlyGlnValSerTrpTyrAlaAsnGlyProIleLeuLysAspGlnLeu 221
QY 598 AGCTTCAGTACGGGAGGAGCAGACACACAGCGCTCAGGGTTCATCGGTCACATCAGTAC 657
Db 222 GlyLeuGlnLeuTrpGlyArgGlyPheThrArgGlyGluAspArgile-----LeuAsn 239
QY 658 GATACAGCAGCAGCGGTATTCCTTATCCACGGAGTCACAGAAATTAATCTTGGTCA 717
Db 240 GlyThrThrGly-----AlaLysGluTyrAspPheAsnGly 251
QY 718 CGTCTTGACTGGAAGCGCTCGAGCAGGATGTGCTCTGTTTGATATGATACACCGCG 777
Db 252 ArgLeuThrPheThrProAsnGluAspHisAspIleTyrLeuGluGlyGlyLysThrArg 271
QY 778 CAGCGTTATGATAACCGGATGGCACTGGGAGTCTCAGCGGGGATATGACCGGACC 837
Db 272 LeuArg-----ArgAspAlaGluProGlyAspThrLeuAlaAlaThrAspAlaAsn 288
QY 838 CTCGCTATGACGGAAC---AAAAATTCAGCTGGCTATGATCATATTCACCTTCGGA 894
Db 289 GlyThrTyrAsnThrAsnThrArgAspHisTrpSerLeuSerHisThrGlyArgTrpGly 308
QY 895 -----ACATGGAAATCGTATCTATCTGAACTGGAACGAG 924
Db 309 ProThrThrSerLeuPheSerPheGlnGluTrpAlaGluArgThrAsnPheThrArg 328
QY 925 ACAGAAATAAAGTCTGAGCTGTGACGAGTGTACTGAGCGGACAAATGGGGGCTT 984
Db 329 AsnThrArgThrGlyArg-----Val 335
QY 985 GCGCGTCAGCGCGGAGCTTAAGGAATCGAACCTTATCTGTAATTCATTCTGCTTACC 1044
Db 336 ThrGluAsnProArgSerProGluIleArgAsnThrValLeuAspGlyLysPheThrThr 355
QY 1045 CCT-----CTGGGAGAAATCTACTCTGGTTACCGTGGGGGGCGAGTTTCAGAGCTCGTCC 1098

Db 356 ProPheGluLeuPheGlyAsnHisThrLeuValThrGlyGlyGlnTyrPheGluAlaArg 375
QY 1099 ATGAAGACGGAGTGTCTCTTCACACACA-----GGTGAACATTTCCGGCAGAAAGC 1152
Db 376 LeuThrAspGlnAsnProGlyArgArgThrAspArgAspGluThrPheSerAlaThrGln 395
QY 1153 TGTCTCGGTATTTCTGAGGATGAGTGGCATCTCAGGATGCATCTGCGTGTGCTGCTGGCGC 1212
Db 396 TrpAlaLeuPheLeuGluAspGluTrpArgIleValAspAsnPheAlaLeuThrGlyGly 415
QY 1213 AGCCGTATGAACATCATCAGCAATTCGGGGGACACTTCAGTCCGCGTGCATATCTGGTC 1272
Db 416 LeuArgLeuAspAsnHisGlnLysTyrGlyAsnHisPheSerProArgLeuTyrGlyVal 435
QY 1273 TGGGATGTGGCAGATCCCTGAGCGTGAAGCGGTGTGACACCGGATATTAAGGACCC 1332
Db 436 TrpSerAlaThrGluGluLeuThrIleLysGlyGlyValSerThrGlyPheArgAlaPro 455
QY 1333 AGAATGGGCGAGCTACATAAGGATTAAGTGTGTCTCGCGGACGAGAAACAAAT--- 1389
Db 456 GluIleArgGlnIleAlaProGlyTyrAlaTyrThrThrGlyGlyGlyCysSerTyr 475
QY 1390 -----CTACTGTGTAAACCCGACCTGAAGCGGAGAGAGC 1425
Db 476 GlyProSerGlyThrCysGlyValIleIleGlyAspProAsnLeuGluAlaGluLysSer 495
QY 1426 GTCAGTTATCAGCGCTGGGTGTATTACGATAACCCCGCTCTGTAATGCAATGTCACA 1485
Db 496 ThrSerTyrGluIleAlaLeuLeuTrpAspAsn---GlyAspIleAlaLeuGlyAlaThr 514
QY 1486 GGTATTATGACTGACTTCTCCAAACAAGATT----- 1515
Db 515 TyrPheTyrThrAspPheLysAspLysIleSerAsnAlaLeuValLeuAsnProAspGly 534
QY 1516 -----GTCCTTATTCATAAAT 1533
Db 535 ThrProAlaArgTrpSerGluAspArgAsnTyrArgLeuTrpTyrAsnTyrAsnIleAsp 554
QY 1534 GATAACACCAATACCTATGTAACAGCGGAAAGCGCGGTGCACCGTGTGGAAATTTGCC 1593
Db 555 Asp-----AlaIleIleGlnGlyValGluLeuThr 564
QY 1594 GGCACATTCGCGCTGTGG-----TCAGAGATGTACGCTGCATCAATTAACCTGGA 1647
Db 565 AlaThr-----TrpTyrAlaThrSerGluLeuThrLeuArgGlyAsnTyrThrTyr 581
QY 1648 ACCGAAAGTGAACACGATGTGTGATACAAAGGTGCGCGCTGAGTTATACCCCTGAA 1707
Db 582 ThrHisSerGluGlnLysThrGlyAspTyrGluGlyPheProLeuAlaArgThrProGlu 601
QY 1708 CACATGTGTGAATCGAAACTCACTGGCAGATCACC-----GAAGAGGTGGCATCATGG 1761
Db 602 HisMetAlaAsnLeuAspGlyAspTrp---ValThrProIleAspGlyLeuGluAlaTrp 620
QY 1762 CTGGGTGCGCGTTATTCGCGGGAACA-----CCAGTTTCACCCAGAAATAT 1809
Db 621 ValSerLeuAsnTyrHisGlySerGluIleAsnAlaGlyProArgIleGlyAlaAsnGly 640
QY 1810 TCCTCAGTACGCGCTGTACAGAAAGAAAGTGTATGATGAGAAAGAGAAATACCTGAAAGCC 1869
Db 641 ThrProValThr-----IleAsnGlyGlnSerGlyArgLysTyrAspAla 655
QY 1870 TGGAGGTGTGTGGATGAGGTCTGTGGAAGATGACGATGCCCTGAGCTGAATGCT 1929
Db 656 TyrThrThrLeuAspIleGlyAlaLysTyrAlaValAlaGluAsnValAspLeuAsnAla 675
QY 1930 GCGGTGAATACCTGCTCAACAAGGATTACAGTACGCTGAGCCTGTACGCGCGGTAAAG 1989
Db 676 AlaValTyrAsnValPheAspLysAsp-----Val 684
QY 1990 AGTACGCTGTATGCGGTGATTACTTCAGACGGGATCATCAACACAGGATATGTGATA 2049


```

Db      460 GlyAlaSerIleMetTyrGlyAsnArgAspLeuLysProGluThrSerValThrGluGlu 479
QY      1438 CTGGGGTGATATACGATAACCCCGCGCTGTAATGCCAATGTACAGGTTTATGACT 1497
Db      480 IleGlyIleIleTyrSerAsnAspSerGlyPheSerAlaSerAlaThrLeuPheAsnThr 499
QY      1498 GACTTCCTCAACAAGATTGTCTCTATTCCATAATGAT----- 1536
Db      500 AspPheLysAsnLysLeuThrSerTyrAspIleGlyThrLysAspProValThrGlyLeu 519
QY      1537 AACACCAATAGCTATGTAACAGCGGAAGCCCGGTGACCGGTGGATTTGCCCGC 1596
Db      520 AsnThrPheIleTyrAspAsnValGlyGluAlaAsnIleArgGlyValGluLeuAlaThr 539
QY      1597 ACATTGCCCTGTGGTCAGAGGATCTCAGCGTGCACCTGATTAACCTGACCTGACCCGAAGT 1656
Db      540 GlnIleProValTyr---AspLysTrpHisValSerAlaAsnTyrThrPheThrAspSer 558
QY      1657 GAACAACGTGATGTGTATAAC-----AAAGTGGCCCGCTGAGT 1695
Db      559 ArgArgLysSerAspAspGluSerLeuAsnGlyLysSerLeuLysGlyGluProLeuGlu 578
QY      1696 TATACCCCTGACACATCGTGAATCGAAAGCTGACCTGGCAGATCACCGAAGAGGTGCA 1755
Db      579 ArgThrProArgHisAlaAlaAsnAlaLysLeuGluIlePheTyrThrGlnAspIleThr 598
QY      1756 TCATGCTGGTGGTGGCCGCTTATCGCGGG----- 1782
Db      599 PheTyrSerSerLeuAsnTyrThrGlyLysGlnIleTrpAlaAlaGlnArgAsnGlyAla 618
QY      1783 AAAACACCAACGTTTCACCCAGAAATTATTCGTCACTGACGCGCTGTACAGAAAGTGTAT 1842
Db      619 LysValProArgValArgAsnGlyPheThrSerMet----- 630
QY      1843 GATGAGAAAGGAGATACCTGAAGCCTGGACGGTGGTGCAGGTCTGTCTGTGGAG 1902
Db      631 -----AspIleGlyLeuAsnTyrGln 637
QY      1903 ATGACGGATGCCCTGACGCTGAATGCTCGGTGAATAACCTGCTCAACAAGGATTACAGT 1962
Db      638 IleLeuProAspThrLeuIleAsnPheAlaValLeuAsnValThrAspArgLysSerGlu 657
QY      1963 GACGTGAGCCTGTACAGTCCCGTAAGAGTACGCTGTATGCCGGTGATTCTCCAGACG 2022
Db      658 AspIleAspThrIleAsp-----GlyAsnTrp----- 666
QY      2023 GGATCATCAACAACAGGATATGTGATACCTGAGCGAAATTAAGTGTGCTGCTGAACAT 2082
Db      667 -----GlnValAspGluGlyArgArgTyrTrpAlaAsnValArgVal 680
QY      2083 CAGTTC 2088
Db      681 SerPhe 682

```

Search completed: October 13, 2004, 11:54:20
Job time : 175 secs

this Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: October 13, 2004, 11:42:27 ; Search time 42 Seconds
(without alignments)

9577.928 Million cell updates/sec

Title: US-10-625-972-4

Perfect score: 3732

Sequence: 1 atggaataaacactctggc.....cgctgaactatcattctga 2091

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -MODEL=frame+...-DEV=xlh

-O=cn2.1/USPTO spool/US10625972/runat 12102004 130106 26831/app_query.fasta_1.2247
-DB=PIR_78 -OPMT=fastan -SUFFIX=trp -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DLOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10625972@CN1_1_54@runat 12102004 130106 26831 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3632	97.3	696	2 H90798	Iha adhesin [impor
2	3632	97.3	696	2 G85607	probable receptor
3	1109	29.7	696	2 D81346	probable iron upta
4	1052.5	28.2	652	2 D82317	iron-regulated out
5	1052.5	28.2	652	2 S25265	outer membrane pro
6	1028	27.5	655	2 AD0160	probable outer mem
7	993	26.6	702	2 AF3038	exogenous ferric s
8	993	26.6	707	2 E98247	outer membrane pro
9	841.5	22.5	663	2 AG0782	colicin I receptor
10	836.5	22.4	659	2 G91009	colicin I receptor
11	832.5	22.3	663	1 QREIC	colicin I receptor
12	827.5	22.2	659	2 A85954	hypothetical prote
13	738	19.8	653	2 F83405	probable TonB-depe
14	629	16.9	746	2 G90706	hypothetical prote

Alignment Scores:
Pred. No.: 7.89e-242 Length: 696
Score: 3632.00 Matches: 695
Percent Similarity: 99.86% Conservative: 0
Best Local Similarity: 99.86% Mismatches: 1
Query Match: 97.32% Indels: 0
DB: 2 Gaps: 0

US-10-625-972-4 (1-2091) x H90798 (1-696)

QY 1 ATGCGAATAACCACTCTGGCTTCGGTAGCTATTCCTCTCGGATTTTCAGCCAGCAGC 60
Db 1 MetArgIleThrThrLeuAlaSerValIleProCysLeuGlyPheSerAlaSer 20
QY 61 ATAGTGTCTGAGAGGATGTGATGCTCTCGGCATCCGCTATGAGAAAAGCTGACT 120
Db 21 IleAlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLeuThr 40

ALIGNMENTS

RESULT 1

H90798 Iha adhesin [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C:Accession: H90798

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: H90798

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-696 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA834783.1; PID:G13360820; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs1360

C:Superfamily: ferrienterochelin receptor, tonB-dependent receptor amino-terminal homolo

15 629 16.9 746 2 B85557
16 625 16.7 746 1 QRCFC
17 604.5 16.2 751 2 AC0574
18 603 16.2 742 2 H83529
19 592.5 15.9 746 2 A40636
20 580.5 15.6 726 2 AG0837
21 541.5 14.5 614 2 A98241
22 541.5 14.5 614 2 F86088
23 528.5 14.2 614 1 QRCBT
24 519 13.9 625 2 AD0476
25 510.5 13.7 614 2 AB0935
26 509.5 13.7 611 2 A82358
27 477 12.8 733 2 A87521
28 448.5 12.0 650 2 B87466
29 440 11.8 539 2 AC3334
30 407 10.9 638 2 AG2219
31 396.5 10.6 616 2 C83385
32 395 10.6 616 2 B83485
33 391 10.5 660 2 D91176
34 391 10.5 660 2 E86022
35 373.5 10.0 700 2 F64897
36 371.5 10.0 700 2 H85732
37 371 9.9 698 2 E82443
38 369.5 9.9 700 2 G90885
39 364.5 9.8 715 2 H90777
40 364.5 9.8 719 2 F85824
41 359.5 9.6 764 2 H83055
42 356 9.5 802 2 AC2075
43 352.5 9.4 713 2 E91118
44 352.5 9.4 713 2 D85863
45 350.5 9.4 686 2 S28042

hypothetical prote
ferrienterochelin
ferrienterochelin
siderophore recept
ferric enterobacti
TonB-dependent out
hypothetical prote
hypothetical prote
vitamin B12 recept
probable vitamin B
vitamin B12 recept
TonB-dependent rec
TonB-dependent rec
metal chelate oute
hypothetical prote
hypothetical prote
probable tonB-depe
heme utilization/c
outer membrane hem
probable tonB-depe
hypothetical prote
heme transport pro
hypothetical prote
hypothetical prote
probable outer mem
outer membrane hem
probable ferrichro
probable iron comp
hemin receptor pre

121 AACGAGCCGCGCTGTTTCTGTTAGTCCAGGAGAAATTCAGTCCAGCCAGTACCAC 180
 41 AsnAlaAlaSerValSerValIleSerGlnGluLeuGlnSerSerGlnTyrHis 60
 181 GATTCGGGGAGGCTCTGAGATCAGTAGAGGGTGTGATGTTGAAAGTGTGACGGGTAA 240
 61 AspLeuAlaGluAlaLeuArgSerValGluGlyValAspValGluSerGlyThrGlyLys 80
 241 ACCGAGGCTGGAAATCAGCATCCGAGGAAATCCAGCCAGTTACAGCTGATACAGT 300
 81 ThrGlyGlyLeuGluIleSerIleArgGlyPheProAlaSerThrLeuIleLeuIle 100
 301 GATGGTTCCTCAGGCGGAGCAGTACGTGACTCCCAACGGTTTTCTGCATGAAT 360
 101 AspGlyValArgGlnGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
 361 ACCGGTTCATGCCCTCTGGCCGCTATTCAGGCTATTCAGGCTATTCAGGGCGCATG 420
 121 ThrGlyPheMetProProLeuAlaAlaIleGluArgIleGluValIleArgGlyProMet 140
 421 TCCACACTGTATGCTCTGATGCGATGGCGGTGTGTGAATATCATTAACCAAGAAAT 480
 141 SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAsnIleIleThrArgLysAsn 160
 481 GCAGACAAATGGCTCTCTTCGTCATGCGGCTGAATCTGAGGCTGAGGAAAGCAATGG 540
 161 AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp 180
 541 GGTAAACAGCAGCAGTAAATTTCTGAGCAGTGGTCCCTCTGATGATTTCTGTCAGC 600
 181 GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspSerValSer 200
 601 CTCAGGTACCGGTAGCACACACAGCGTCAGGGTTTCATCGTCCATCATCTGAGCGAT 660
 201 LeuGlnValArgGlySerThrGlnArgGlnGlySerSerValThrSerLeuSerAsp 220
 661 ACAGCAGCAGCGTATTCCTTATCCCGAGTGTACAGAGTACAGAAATTAATCTTGGTGCAGT 720
 221 ThrAlaAlaThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg 240
 721 CTTCAGTGAAGCGTCGAGCAGGATGTCTCTGTTTCATATGATACCAACCGCGCAG 780
 241 LeuAspTrpLysAlaSerGluGlnAspValLeuTrpPheAspMetAspThrThrArgGln 260
 781 CGTTATGATACCGGATGGGCACTGGGAGTCTGACGGGGGATATACCGGACCTG 840
 261 ArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyGlyTyrAspArgThrLeu 280
 841 CGCTATGAGCGAAACAAAATTTTCAGCTGGCTATGATCATACTTTTCACCTTCGGAACATGG 900
 281 ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspHisThrPheThrPheGlyThrTrp 300
 901 AATCTGATCTGAATCGGACGACGACAGAAAATAAAGTCTGTCAGCTTGTACGCGAGTGA 960
 301 LysSerTyrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal 320
 961 CTGAAGCGGCAAAATGGGGCTTGGCTGAGCGCGGAGCTTAAGGAATCGAACCTT 1020
 321 LeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu 340
 1021 ATCCTGAATTCATCTGTTACCTCTGGAGAAATCTCATCTGTTACGTTACGGTGGGGGC 1080
 341 IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly 360
 1081 GAGTTTCAGAGCTCGTCCATGAAGACGAGTGTCTTCGCGACGACAGGTTGAATCTTC 1140
 361 GluPheGlnSerSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe 380
 1141 CGGCAGAAAGCTGGTGGTATTTGCTGAGGATGAGTGCATCTCACGAGTGCATCTGCG 1200
 381 ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspAlaLeuAla 400

1201 CTGACTGCGGCGAGCCGCTATGAACATCATGAGCAATTCGGGGGACACTTCAGTCCCGT 1260
 401 LeuThrAlaGlySerArgTyrGluHisGluGlnPheGlyGlyHisPheSerProArg 420
 1261 GCATATCTGGTCTGGATGTGGCAGATGCTCGTGGACGCTGAAAGCGGTGTGACCCAGGA 1320
 421 AlaTyrLeuValTrpAspValAlaAspAlaTrpThrLeuLysGlyValThrThrGly 440
 1321 TATAAGGCACCCAGAAATGGGCGAGCTACATAAAGGATTAAGTGTGTGTCGGGCGAGGA 1380
 441 TyrLysAlaProArgMetGlyGlnLeuHisLysGlyIleSerGlyValSerGlyGlnGly 460
 1381 AAAACAAATCTACTTGTGTACCCCGACTGAAGCGGAGAGAGCGTCAAGTTATGAGCT 1440
 461 LysThrAsnLeuLeuGlyAsnProAspLysProGluGluSerValSerTyrGluAla 480
 1441 GGGGTGATTAACGATAACCCCGCGTCTGAATCCCAATGTCACAGGTTTATGACTGAC 1500
 481 GlyValTyrTyrAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp 500
 1501 TTCTCCAAACAGATGTCTCTTATTCATATAATGATAACCAATAGCTATGTAACAGC 1560
 501 PheSerAsnLysIleValSerTyrSerIleAsnAspAsnThrAsnSerTyrValAsnSer 520
 1561 GGAAGGCCCGGTTCAGCGGTGGGAATTTGCCGGCACATTCGGCTGTGTCAGGAGAT 1620
 521 GlyLysAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp 540
 1621 GTCACGCTGTCACTGAATTAACCTGACCCGAGTGAACCAACGTGATGTTGATAACAAA 1680
 541 ValThrLeuSerLeuAsnTyrThrTrpThrArgSerGluGlnArgAspGlyAspAsnLys 560
 1681 GGTGCGCGCTGAGTTATACCCCTGAACACATGTTGAATCGGAAACCTGAACCTGAGATC 1740
 561 GlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle 580
 1741 ACCGAAGAGTGGCATCATGCTGGTGGTCCGCTTATCGCGGAAACACACACCTTTTCAAC 1800
 581 ThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgGlyLysThrProArgPheThr 600
 1801 CAGAATTTATGTCACCTGACGCTGTACAGAAAGTGTATGATGAGAAAGAGAGATAC 1860
 601 GlnAsnTyrSerSerLeuSerAlaValGlnLysLysValTyrAspGluLysGlyGluTyr 620
 1861 CTGAAGCCCTGGAGCGGTGGATGCGAGTCTCGTGGAGATGACGGATGCGCTGACG 1920
 621 LeuLysAlaTrpThrValValAspAlaGlyLeuSerTrpLysMetThrAspAlaLeuThr 640
 1921 CTGAATCTCGGTGAATAACCTGCTCAACAGGATTACAGTACGCTGAGCCTGTACAGT 1980
 641 LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyrSerAspValSerLeuTyrSer 660
 1981 CGCGTGAAGTAGCTGTATGCGGTGATTCCTTCAGACGGGATCATCAACACAGGA 2040
 661 AlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerThrThrGly 680
 2041 TATGTGATCTCAGCGCAAAATTAAGTGTGCTCGCTCAACTATCATCAGTTC 2088
 681 TyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe 696

RESULT 2
 G85607
 probable receptor Z1178 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-feb-2001 #text_change 14-Sep-2001
 C:Accession: G85607; H95658
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; PMID:21074935; PMID:11206551
 A:Accession: G85607
 A:Status: preliminary

A:Molecule type: DNA	
A:Residues: 1-696 <ST>	
A:CROSS-references: GB:AB005174; NID:G12513986; PIDN:AAG55323.1; GSPDB:GN00145; UWGP:Z11	
A:Experimental source: strain O157:H7, substrain EDL933	
A:Accession: H85658	
A>Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-696 <ST>	
A:CROSS-references: GB:AB005174; NID:G12514503; PIDN:AAG55732.1; GSPDB:GN00145; UWGP:Z16	
A:Experimental source: strain O157:H7, substrain EDL933	
C:Genetics:	
A:Gene: Z1178; Z1617	
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolog	
Alignment Scores:	
Pred. No.:	7,896-242
Score:	2632.00
Percent Similarity:	99.86%
Best Local Similarity:	99.86%
Query Match:	97.32%
DB:	2
US-10-625-972-4 (1-2091) x G85607 (1-696)	
QY	1 ATGCGAATAACACACTCTGGCTTCCTAGTATCCCTGTCGGATTTTCAGCCAGCAGC 60
DB	1 MetArgIleThrLeuAlaSerValValIleProCysLeuGlyPheSerAlaSerSer 20
QY	61 ATAGCTGCTGCAGAGGATGTGATGTCCTCGCATCCGGCTATGAGAAAGCTGACT 120
DB	21 IieAlaAlaAlaGluAspValMetIleValSerAlaSerGlyTyrGluLysLeuThr 40
QY	121 AACGAGCGCGCAGTGTCTCTGTATAGCCAGAGGAGATTCAGTCAGTCAGCCAGTACCAC 180
DB	41 AsnAlaAlaAlaSerValSerValIleSerGlnGluGluLeuGlnSerSerGlnTyrHis 60
QY	181 GATCTGGCGGAGCTCTGAGATCAGTACGAGGTGGATGTTGAAGTGGTACGGGTAAA 240
DB	61 AspLeuAlaGluAlaLeuArgSerValGluGluValAspValGluSerGlyThrGlyLys 80
QY	241 ACCGAGGGCTGAAATCAGCATCCGAGGAATGCCAGCGATTACACCTGATCTGATT 300
DB	81 ThrGlyGlyLeuGluIleSerIleArgGlyMetProAlaSerTyrThrLeuIleLeuIle 100
QY	301 GATGTTGTCGTCAGGGCGAGCAGTACGTCGACTCCCAACGGTTTTCGCCATGAT 360
DB	101 AspGlyValArgGlnGlySerSerAspValThrProAsnGlyPheSerAlaMetAsn 120
QY	361 ACCGGTTTCATGCCCTCTGCGCCATTGAGCGTATTGAGCTTATCAGGGGGCGGATG 420
DB	121 ThrGlyPheMetProProLeuAlaAlaIleGluArgIleGluValIleArgGlyProMet 140
QY	421 TCACACTGTATGGCTCTGATCGGATGGCGGTGGTGAATATCATACCGAAGAAAT 480
DB	141 SerThrLeuTyrGlySerAspAlaMetGlyGlyValValAlaSerIleIleThrArgLysAsn 160
QY	481 CGAGCAAAATGGCTCTCTTCGTCATGTCAGGGCTGAATCTCAGGAAGCAACAAATGG 540
DB	161 AlaAspLysTrpLeuSerSerValAsnAlaGlyLeuAsnLeuGlnGluSerAsnLysTrp 180
QY	541 GTPAACAGCAGCCAGTTTAAATTCGTGAGCAGTGGTCCCTTTGTGGATGATCTGTGACG 600
DB	181 GlyAsnSerSerGlnPheAsnPheTrpSerSerGlyProLeuValAspAspSerValSer 200
QY	601 CTCAGGTACGGGTAGCACACACAGCTCAGGTTCATCGGTTCATCATCTGAGCGCAT 660
DB	201 LeuGlnValArgGlySerThrGlnGlnArgGlnGlySerSerValThrSerLeuSerAsp 220
QY	661 ACAGCAGGACGGCTATCTCTATCCACGGAGTCACAGAATTAAATCTTGTGTCACGT 720
DB	221 ThrAlaAlaThrArgIleProTyrProThrGluSerGlnAsnTyrAsnLeuGlyAlaArg 240
QY	721 CTTGACTGGAGCGCTCGGAGCAGGATGTCTCTGTTGATATGGATACCCCGGCAG 780

Db	241 LeuAspTrpLysAlaSerGluGlnAspValLeuTrpPheAspMetAspThrThrArgGln 260
QY	781 CGTTATGATAAACCAGGATCGCAACTGGGGAGTCTGACGGGGGGATATACCGGACCTG 840
Db	261 ArgTyrAspAsnArgAspGlyGlnLeuGlySerLeuThrGlyTyrAspArgThrLeu 280
QY	841 CGCTATGAGCGAAACAAATTCAGTGGCTGTATGATCATCTTACCTTCGGAACATGG 900
Db	281 ArgTyrGluArgAsnLysIleSerAlaGlyTyrAspHisThrPheThrPheGlyThrTrp 300
QY	901 AAATCGTATCTGAACCTGGACGACAGAGAAATAAAGGTCTGTCAGCTTGTACCGACTGTA 960
Db	301 LysSerTyrLeuAsnTrpAsnGluThrGluAsnLysGlyArgGluLeuValArgSerVal 320
QY	961 CTGAAGCGGACAAATGGGGCTTGGCGGTAGCCCGGAGCTTAAAGAACTGAACCTT 1020
Db	321 LeuLysArgAspLysTrpGlyLeuAlaGlyGlnProArgGluLeuLysGluSerAsnLeu 340
QY	1021 ATCTGAAATTCATCTGCTTACCTCTGGGAGATCTCATCTGTTACGGTGGGGGC 1080
Db	341 IleLeuAsnSerLeuLeuLeuThrProLeuGlyGluSerHisLeuValThrValGlyGly 360
QY	1081 GAGTTTCAGAGCTCGTCCATGAAAGACGAGGATTTGCTTTGCCACGACAGGTGAACCTTC 1140
Db	361 GluPheGlnSerSerSerMetLysAspGlyValValLeuAlaSerThrGlyGluThrPhe 380
QY	1141 CGGCAAAAAGCTGGTGGTATTGTCGAGGATGAGTGGCATCTCAGGATGACACTGGC 1200
Db	381 ArgGlnLysSerTrpSerValPheAlaGluAspGluTrpHisLeuThrAspAlaLeuAla 400
QY	1201 CTGACTCGGGCAGCGCTATGAACATCATGACCAATTCGGGGGACACTTCAGTCCGCGT 1260
Db	401 LeuThrAlaGlySerArgTyrGluHisHisGluGlnPheGlyGlyHisPheSerProArg 420
QY	1261 GCATATCTGTCGGATGGGAGATGCGTGGAGCTGAAAGCGGTGTGACACGGGA 1320
Db	421 AlaTyrLeuValTrpAspValAlaAspAlaTrpThrLeuLysGlyGlyValThrThrGly 440
QY	1321 TATTAAGCACCCGAAATGGGGCAGCTACATAAAGGATATTAGTGGTGTCTCCGGCAGGA 1380
Db	441 TyrLysAlaProArgMetGlyGlnLeuHisLysGlyLysSerGlyValSerGlyGlnGly 460
QY	1381 AAAACAAATCTACTTGGTAAACCCCGCTGAAAGCGGAGAGAGCGTCAGTTATGAGCT 1440
Db	461 LysThrAsnLeuLeuGlyAsnProAspLeuLysProGluGluSerValSerTyrGluAla 480
QY	1441 GGGGTGTATTACGATAACCCCGCGCTGCAATGCCCAATGTCACAGGTTTTATGACTGAC 1500
Db	481 GlyValTyrTyrAspAsnProAlaGlyLeuAsnAlaAsnValThrGlyPheMetThrAsp 500
QY	1501 TTCTCCCAACAGATTGTCTCTTATTCATAAAATGATAACCAATAGCTATGTAACAGC 1560
Db	501 PheSerAsnLysIleValSerTyrSerIleAsnAspAsnThrAsnSerTyrValAsnSer 520
QY	1561 GGAAGAGCCCGGTTCACGGTGTGAATTTGCCGGCATTGCCGCTGTGGTTCAGAGAT 1620
Db	521 GlyLysAlaArgLeuHisGlyValGluPheAlaGlyThrLeuProLeuTrpSerGluAsp 540
QY	1621 GTCAAGCTGTACCTGAATTAACCTGCACCGAGTCAACAACCTGATGGTGAACAAA 1680
Db	541 ValThrLeuSerLeuAsnTyrThrTrpThrArgSerGluGlnArgAspGlyAspAsnLys 560
QY	1681 GGTGGCGCTGTAGTTATACCCCTGAAACACATGGTGAATGCGAACTGAACCTGCAGATC 1740
Db	561 GlyAlaProLeuSerTyrThrProGluHisMetValAsnAlaLysLeuAsnTrpGlnIle 580
QY	1741 ACCGAAGGTGGCATCATGGCTGGTGGCTGGTTCATCGGGGAAACACACGCTTTCACC 1800
Db	581 ThrGluGluValAlaSerTrpLeuGlyAlaArgTyrArgGlyLysThrProArgPheThr 600
QY	1801 CAGAAATTATTCCTCCTCAGCGCTGTACAGAAAGTGTATGATGTAAGAGGAGATAC 1860

Db 601 GlnAsnTyrSerSerLeuSerAlaValGlnLysValTyrAspGluLysGlyGluTyr 620
 QY 1861 CTGAAGCGCTGGACGGTGGTGGATGACAGGCTGTGCTGGGAAGATGACGATGCCCTGACG 1920
 Db 621 LeuLysAlaThrThrValValAspAlaGlyLeuSerTrpLysMetThrAspAlaLeuThr 640
 QY 1921 CTGAATCGCTGGTGAATACCTGCTCAACAGATTACAGTGACGTGACCTGTACAGT 1980
 Db 641 LeuAsnAlaAlaValAsnAsnLeuLeuAsnLysAspTyrSerValSerLeuTyrSer 660
 QY 1981 GCCCGTAAGAGTACCTGTATGCGGTGATTACTTCCAGACGGATCATCAACACAGCA 2040
 Db 661 AlaGlyLysSerThrLeuTyrAlaGlyAspTyrPheGlnThrGlySerSerThrThrGly 680
 QY 2041 TATGTGATACCTGAGCAATTAATCTGGATGCTGCTGAATCATCATCAGTTC 2089
 Db 681 TyrValIleProGluArgAsnTyrTrpMetSerLeuAsnTyrGlnPhe 696
 RESULT 3
 D81346
 Probable iron uptake protein Cj0755 [imported] - Campylobacter jejuni (strain NCTC 11168)
 C/Species: Campylobacter jejuni
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C/Accession: D81346
 R/Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell
 Nature 403, 665-668, 2000
 A/Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
 A/Reference number: A81250; MUID:20150912; PMID:10688204
 A/Accession: D81346
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-696 <PAR>
 A/Cross-references: GB:AL139076; GB:AL111168; NID:96968128; PIDN:CAB73021.1; PID:9696820
 A/Experimental source: serotype O2, strain NCTC 11168
 C/Genetics:
 C/Supfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolo

Alignment Scores:
 Pred. No.: 2,29e-68 Length: 696
 Score: 1109.00 Matches: 255
 Percent Similarity: 54.80% Conservative: 150
 Best Local Similarity: 34.51% Mismatches: 248
 Query Match: 29.72% Indels: 86
 DB: 2 Gaps: 16

US-10-625-972-4 (1-2091) x D81346 (1-696)

QY 1 ATCGGAATACCACTCTGGCTCCGAGTTCATTCCTCGATTCCTCGATTTTCAGCCAGCAGC 60
 Db 1 MetLysLysIleCysLeuSerValCysAlaIleGlyLeuLeuSerSerAsnAlaIleSer 20
 QY 61 ATAGCTGCTGCAGAGATGTGATGTTCTCGGATCCGGTATCAGAAAAGCTGACT 120
 Db 21 GlnAsnValGluLeuAspSerSerIleValSerAlaSerGlyPheThrGlnAspIleLys 40
 QY 121 AACCGACCGCCAGTGTCTTCTGTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCAC 180
 Db 41 GluAlaProAlaThrIleAsnValIleThrLysLysGluLeuGlnSerLysProTyrArg 60
 QY 181 GATCTCGCGGAGCTCTGAGATCAGTAGAGGTGTGGATGTTGAAAGTGGTACGGGTAAA 240
 Db 61 AspValAlaGluAlaIleAlaAspIleProGlyValAspLeuTyrAlaSerLysGlyLys 80
 QY 241 ACCGGAGGGCTGGAAATCAGCATCCGAGGAATCCAGCCAGGTACACGCTGACTACTGATT 300
 Db 81 ThrGlySerTyrAsnIleThrMetArgGlyIle---ThrGlyTyrThrLeuValLeuIle 99
 QY 301 GATGTGTTCGTCAGCGCGGAGACAGTACGTACGTACCTCCCAAGCGTTTCTGCCCATGAAT 360
 Db 100 AspGlyArgArgGlnGlyIleGlyGlyGluValGlyProAsnGlyPheAsnGluIleSer 119

QY 361 ACCGGTTTCATGCCCCCTCTGGCCGCCATTGAGCCTATTGAGTTATCAGGGGCGCGATG 420
 Db 120 AsnSerPheLeuProIleSerSerIleGluArgIleGluValIleLysGlyProMet 139
 QY 421 TCCACACTGATGCTCTGATGCGATGGCGGTGCTGGTGAATATCATACACAGAAAGAT 480
 Db 140 SerThrLeuTyrGlySerGluAlaLeuGlyValValAsnIleIleThrLysLysVal 159
 QY 481 GCAGCAAAATCGCTCTCTCCGTCAATCGACGGGTGAATCTGCAGGAAGACCAAAATGG 540
 Db 160 SerAspLysTrpGluThrSerValSerLeuAspAlaLeuLeuAsnGluAsnLysAspTrp 179
 QY 541 GGTAAACAGCAGCAGTAAATTTCTGGAGCAGTGTCCCTTGTGGATGATTCTGTGTCAGC 600
 Db 180 GlyAsnThrTyrGlyThrSerIleTyrSerSerGlyProLeuMetAsnAspLysLeuGly 199
 QY 601 CTGCAGGTACGCGGTAGCACACACACGCTCAGGGTTCATCGGTGCACATCCTGAGCGAT 660
 Db 200 LeuThrLeuArgPheArgGluPheTyrArgGlnGlnSerAsnVal---GluPheThrAsn 218
 QY 661 ACAGCAGCAGCGCTATT-----CCTTATCCACGGAGTGCACAGAAATAT 705
 Db 219 GlySerGlyGlnArgValGlnGlyAspGlnAlaGlnSerProThrLysAlaAsnAspPhe 238
 QY 706 AATCTTGTGTCACGCTCTTGAAGCGGTGCGGAGCAGGATGTGCTCTGGTTTGATATG 765
 Db 239 AsnIleGlyThrArgIleSerTyrLeuAlaAsnAspTyrAsnThrPheIlePheAspIle 288
 QY 766 GATACCAACCGCGCAGCGCTTATGATAACCGGGATGGGCAACTG----- 807
 Db 259 AspPheSerArgAsnHisTyrAspAsnLysGlnGlyGlnLeuGlyThrIleThrSerPro 278
 QY 808 -----GGAGCTCTCAGCGGGGATATCACCGGACCTCGCTATGAGCGAAAC 855
 Db 279 GlyArgThrProGlySerLeuThrGlyGlyTyrAlaAspIleMetGluValAspLysPhe 298
 QY 856 AAAATTTTCAGCTGCTATGATCATCTTTCACCTTCGGAACATCGAAATCGTATCTGAAC 915
 Db 299 ValThrTyrLeuSerHisGluGlyValTyrGluAsnPheSerIleThrSerGlyLeuGln 318
 QY 916 TGGAAACGAGACAGAAATAAGTCTGACCTGTGACGCTGTACGAGCGCGACAAA 975
 Db 319 TyrAsnArgValSerAsnAspGlyArgGluValVal----- 330
 QY 976 TGGGGGCTTCCGCTCAGCCG-----CGGAGCTTAAGGAATCGAACCTT 1020
 Db 331 ---GlyGlnSerThrGlnProPheLeuGlyGluAsnArgAspIleValAlaGluAspIle 349
 QY 1021 ATCTGTAATTCATTCATCTTACCTCTCGGAGATCTCATCTGTTACGGTGGGGGGC 1080
 Db 350 IleLeuAspThrLysSerValIleProLeuGlyGlnSerHisIleLeuSerValGlyGly 369
 QY 1081 GAGTTTCAGAGCTCGTCCATGAAGAGCGGAGTGTCTTCCACGACAGGTGAACCTTTC 1140
 Db 370 GluTyrArgLeuGluLysMetGlnAspLysIle-----AlaSerProThrAsnPhe 386
 QY 1141 CGGCAAAAAGCTGGTGGTATTGCTGAGGATGAGTGGCATCTCAGGATGCATTCGG 1200
 Db 387 AspGlnTyrLeuLeuAlaIlePheAlaGluAspGluTyrSerIleLysAspAspLeuArg 406
 QY 1201 CTGACTCGCGGAGCGCTATGAACATCATGACCAATTCGGGGGACACTTCATCGCGCT 1260
 Db 407 LeuThrPheGlyAlaArgTyrAsnHisHisGluIlePheGlyAsnAsnValSerProArg 426
 QY 1261 GCATATCTGCTCGGATGTGGCAGATGCTGCAGCTGAAAGCGGTGTGACCGGGA 1320
 Db 427 AlaTyrValValTyrAsnProThrAsnGluLeuThrLeuLysGlyGlyValSerThrGly 446
 QY 1321 TATAAGCCACCCAGATGGGCGAGCTACATAAGGATTAGTGGTGTCTCCGGGAGGGA 1380
 Db 447 PheArgThrProTyrAlaAsnArgLeuIleAsnGlyThrTyrSerTyrSerGlyGlnGly 466
 QY 1381 AAAACAAATCTACTTGGTAAACCCCGACCTGAAGCCCGAGAGAGCGCTCAGTTATGAGCT 1440


```

...
467 ArgPheProThrTyrGlyAsnProAspLeuLysGluThrSerLeuAsnTyrGluLe 486
1441 GGGGTGTTACGATAACCCCGCGTCTGAATGCAATGTCACAGGTTTATGACTGAC 1500
487 AlaAlaIleTyrAsnAsnAspLeu---PheTyrValSerAlaThrGlyPheLeuThrAsn 505
1501 TTCTCCAAACAGATTGCTCTTATTCATAAT----- 1533
506 PheLysAspLysIleSerSerGlnSerTyrAsnSerGluProIleProGlyIleGly 525
1534 -----GATAACACCAATAGCTATTAACACGGGAAGCCCGCGTTCACGGT 1581
526 ThrCysAspAlaAspArgCysSerAlaIleAsnHisGlyLeValGluTyrLysGly 545
1582 GTGGGAATTT---GCCGGCACAATGCGCTGTGTCAGAGGATGTCACGCTCTACTGAAT 1638
546 ValGluLeuGlyAlaGlyIleSerProLeu-----AspAsnLeuAsnValAsnPheAla 563
1639 TACACTGGACCCGGAAGTGAACACAGTGAATGATGTGATAACAA-----GGTCCGCG 1689
564 TyrThrTyrLeuAspThrGluValLysGluAlaGlnAspArgSerValIleGlyLysPro 583
1690 CTGAGTTATACCCCTGAACACATGTGAATGCGAACTGAATGCGCAGATCACCGAAGAG 1749
584 GluGlnAspSerLeuLysHisAsnIleMetLeuLysThrGluTyrSerPheTyrAsnLys 603
1750 GTGGCATCATGCTGCGTGCC-----CGTTATCGCGGGAACACCCAGT 1794
604 IleThrProTyrLysGlyGluTyrGlnIleAspArgTyrMetGlyAspThr----- 621
1795 TTCACCCAGAAATATTCGTCACCTGAGCGGTGTACAGAGAAAGTGTATGATGAGAGAAG 1854
622 -----AsnIleAsnArg 625
1855 GAATACCTGAAGCCGTGGAGTGTGGATGAGTCTGTGTTGGAAGATGACCGATGCC 1914
626 GluTyrTyrLysAspIlePheLeuAlaSerMetGlyValArgTyrAspIleAsnLysGln 645
1915 CTGACGCTGAATGCTGCGGTGAATACCTGCTCAACAGGATTCAGTAC---GTGAGC 1971
646 TTPSerIleAsnAlaIleTyrAsnLeuPheAspLysSerPheThrAsnGlyTyrGlu 665
1972 CTGTACAGTCCGCGTAAGAGTACGCTGTATGCGGTGATTCCTCCAGACGGGATCATCA 2031
666 SerTyrAlaSerGlySerGlySerThrTyrPvalAsnThrTyr----- 679
2032 ACAACAGGATATGTGATACCTGAGCGAATTAATGATGCTGCTGCACTACGATTC 2088
680 -----AsnArgIleGluGluGlyArgArgMetTyrIleSerIleAsnGlyAsnPhe 696

RESULT 4
D82317
iron-regulated outer membrane virulence protein, TonB receptor family VC0475 [imported]
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: D82317
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: D82317
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-652 <HEI>
A:Cross-references: GB:AE004134; GB:AE003852; NID:9954900; PIDN:AAF93648.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0475
A:Map position: 1
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolo

```

```

Alignment Scores:
Pred. No.: 1.74e-64 Length: 652
Score: 1052.50 Matches: 258
Percent Similarity: 51.80% Conservative: 116
Best Local Similarity: 35.73% Mismatches: 223
Query Match: 28.20% Indels: 125
DB: 2 Gaps: 24

US-10-625-972-4 (1-2091) x D82317 (1-652)
QY 40 CTCGGATTTTCAGCCAGCAGATAGCT-----GCTCAGAGATGTGATGATT 87
DB 17 LeuMetPheSerAlaSerAlaPheAlaGlnAspAlaThrLysThrAspGluThrMetVal 36
QY 88 GTCTCGCATCCGCTATGAGAAAAGCTGACTAACGCCAGCCCGCTGTTCTGTGATT 147
DB 37 ValThrAlaAlaGlyTyrAlaGlnValIleGlnAsnAlaProAlaSerIleSerValIle 56
QY 148 AGCCAGAGAAATGGAGTCCAGCCAGTACACAGTCTGCGGAGGCTGTGAGATCAGTA 207
DB 57 SerArgGluAspLeuGluSerArgTyrTyrArgAspValThrAspAlaLeuLysSerVal 76
QY 208 GAGGGTGTGATGTTGAAAGTGTACGGGTAAACCGGAGGCTGGAATTCAGATCCGA 267
DB 77 ProGlyValThrValThrGlyGlyAspThrThr-----AspIleSerIleArg 93
QY 268 GGAATGCCAGCCAGTTACAGCTGATCTGATGATGTTGTTGTCAGGCGGAGAGCAGT 327
DB 94 GlyMetGlySerAsnTyrThrLeuIleValAspGlyLysArgGln---ThrSerArg 112
QY 328 GACGTGACTCCCAACGGTTTTCT---GCCATGAATACCGGTTTCATGCCCTCTCGCC 384
DB 113 GlnThrArgProAsnSerAspGlyProGlyIleGluGlnGlyTyrLeuProLeuGln 132
QY 385 GCCATTGAGGCTATTAGGTTATCAGGCGCCGATGTCACACTGATGCTCTGTGATGCG 444
DB 133 AlaIleGluArgIleGluValIleArgGlyProMetSerThrLeuTyrGlySerAspAla 152
QY 445 ATGGCGGTGTGGTGAATATCATTACAGAAAGATGCAGACAAATCGCTCTCTTCGTC 504
DB 153 IleGlyGlyValIleAsnIleIleThrArgLysAspGlnGlnTyrSerGlyAsnVal 172
QY 505 AATCCAGGCTGAATCTCGAGGAAGCAACAATGGGGTAACACGACCGCTTAAATTC 564
DB 173 GlnLeuSerThrValValGlnGluAsnArgAlaSerGlyAspGluGlnSerAlaAsnPhe 192
QY 565 TGGAGCAGTGTCCCTTGTGATGATCTCTGAGCTGAGGTACGCGGTAGCACACAA 624
DB 193 PheValThrGlyProLeu---SerAspAlaLeuSerLeuGlnValTyrGlyThr 211
QY 625 CAGCGTTCAG-----GGTTCATCGTCACATCATCTGAGC 657
DB 212 GlnArgAspGluAspGluIleGluHisGlyTyrGlyAspLysSerLeuArgSerLeu--- 230
QY 658 GATACAGCAGCCAGCGTATCTCTATCCACGAGTACAGAGATTATCTTGGT--- 714
DB 231 -----ThrSerLysLeuAsnTyrGlnLeuAsnPro 240
QY 715 -----GCACGCTTGTACTGGAAGCGCTGCGAGCAGGATGTGCTCTGTTTGTATG 765
DB 241 AspHisGlnLeuGlnLeuGluAlaGlyValSerAlaGlnAsp----- 254
QY 766 GATACACCCCGCAGCGTTATGATAACCGGATGGCACTGGGAGTCTGACGGGGGA 825
DB 255 -----ArgGluAsnAsnValGlyLysSerAlaGlnSerSerGlyCysArgGlyThr 271
QY 826 TATGACCGGACC---CTCGCTGTAGCGGAACAAATTTTCAGCTGGCTATGATCATACT 882
DB 272 CysSerAsnThrAspAsnGlnTyrArgArgAsnHisValAlaValSerHisGln----- 289
QY 883 TTCACCTTCGGAACATGGAAA-----TCGTATCTGAATCGAACGACG 924
DB 113

```


290 -----GlyAspTrpGlnAspValGlyGlnSerAspThrTrpLeuGlnTrpGluGlu 306
 925 ACAGAAATAAGGTCGTGAGCTTGACGAGTGTACTGAGCGGACAAATGGGGCTT 984
 307 AsnThrAsnLysSer----- 311
 985 GCCGGTCAGCGCGGAGCTTAAGAAATCGAACTTATCCTGAATTCATTACTGCTTACC 1044
 312 -----ArgGluMetSerIleAspAsnThrValPheLysSerThrLeuValAla 327
 1045 CCTCTGGGAGATCTCATCTGCTTACGCTGGGGGGGAGTTTCAGAGCTGCTCATGAAA 1104
 328 ProIleGlyGlu---HisMetLeuSerPheGlyValGluGlyLysHisGluSerLeuGlu 346
 1105 GACGAGTGTCTCTGTCGACGACAGGTAACCT---TTCGGGAGAGAAAGCTGTCGGTA 1161
 347 AspLysThrSerAsnLysIleSerSerArgThrHisIleSerAsnThrGlnTrpAlaGly 366
 1162 TTCTCTGAGATGAGTGGCATCTCAGGATGCACTTGGCTGACTCGGGGACGCGCTAT 1221
 367 PheIleGluAspGluTrpAlaLeuAlaGluGlnPheArgLeuThrPheGlyGlyArgLeu 386
 1222 GAACATCATCAAGAAATTCGGGGGACACTTCAGTCCGCGTGCATATCTGCTGGGATG 1281
 387 AspHisAspLysAsnTrpGlySerHisPheSerProArgValTrpGlyValTrpAsnLeu 406
 1282 GCAGATCCCTGAGCGCTGAAGCGGTGTGACCGAGGATATAAGACCCAGAAATGGGG 1341
 407 AspProLeuTrpThrValLysGlyValSerThrGlyPheArgAlaProGlnLeuArg 426
 1342 CAGCTACATAAGGGATTAGTGTGTGTCGGGACGAGGAAACAAATCTACTTGGTAA 1401
 427 GluValThrProAspTrpGlyGlnValSerGlyGlyGly-----AsnIleTrpGlyAsn 444
 1402 CCCGACCTGAAGCGGAGAGCGCTGATGAGCTGGGTGTATACGATTAACCC 1461
 445 ProAspLeuLysProGluThrSerIleAsnLysGluLeuSerLeuMetTrpSerThrGly 464
 1462 GCCGCTGTAATGCCAATGTCAGAGTTTATGACTGACTTCTCCAAAGATTTGCTCT 1521
 465 SerGlyLeuAlaAlaSerLeuThrAlaPheHisAsnAspPheLysAspLysIleThrArg 484
 1522 TATTCATAAATGATAAC-----ACCAATAGCTAT----- 1551
 485 ValAlaCysProAlaAsnIleCysThrAlaGlyProAsnGlnTrpGlyAlaAlaProThr 504
 1552 -----GTAACAGCGGAAAGCGGTTGCGAGGTGGAATTTCCGCGACATTCGCG 1605
 505 TyrArgValAsnIleAspGluAlaGluThrTrpGlyAlaGluAlaThrLeuSerLeuPro 524
 1606 CTGTGTCAGAGCATGTCAGCTGTCACCTGAATACACCTGGACCGAGTGAACAAAGT 1665
 525 Ile---ThrGluSerValGluLeuSerSerTrpTrpTrpHisSerGluGlnLys 543
 1666 GATGGTGAATAACAGGTGGCGCTGAGTTATACCCCTGAACACATGGTGAATCGGAAA 1725
 544 SerGlyAsnPheAlaGlyArgProLeuLeuGlnLeuProLysHisLeuPheAsnAlaAsn 563
 1726 CTGAATGCGAGATCAACGAGAGTGGCATATGCTGGTGGCGCTATCCGCGGAAA 1785
 564 LeuSerTrpGlnThrThrAspArgLeuAsnSerTrpAlaAsnLeuAsnTrpArgGlyLys 583
 1786 ACACACGTTTCCACCCAGAAATTTATTCGTCACTGAGCGCTGACAGAAAGATGTATGAT 1845
 584 -----GluMetGlnProGluGlyGlyAla 591
 1846 GAGAAAGGAGATACCTG---AAAGCTCGAGCGGTGGGATGAGCTGCTGCTGGAG 1902
 592 SerAsnAspAspPheIleAlaProSerTrpTrpPheIleAspThrGlyValThrTrpAla 611
 1903 ATGACGAGTGCCTGACGCTGATGTCGGGTGAATAACCTCTCAACAAAGATTAACGT 1962
 612 LeuThrAspThrAlaThrIleLysAlaAlaValTyrAsnLeuPheAspGlnGluValAsn 631

QY 1963 GACGTGAGCCTGTACAGTCGCGGTAAAGAGTAGCTGTATCGCGGTGATTACTTCACG
 Db 632 -----TyrAla---GluTyr----- 635
 QY 2023 GGATCATCAACAACAGGATATGTGATACCTGAGCGAAATTTACTGATGTCGCTGAACAT 2082
 Db 636 -----GlyTyrValGluAspGlyArgGlyTrpLeuGlyLeuAspIle 650
 QY 2083 CAGTTC 2088
 Db 651 AlaPhe 652
 RESULT 5
 S25265
 outer membrane protein irga precursor - Vibrio cholerae
 N:Alternate names: ferrienterochelin receptor homolog
 C:Species: Vibrio cholerae
 C:Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 08-Oct-1999
 C:Accession: S25265; A37834
 R:Goldberg, M.B.; Boyko, S.A.; Butterton, J.R.; Stoeber, J.A.; Payne, S.M.; Calderwood,
 Mol. Microbiol. 6, 2407-2418, 1992
 A:Title: Characterization of a Vibrio cholerae virulence factor homologous to the family
 A:Reference number: S25265; MUID:93023868; PMID:1406279
 A:Accession: S25285
 A:Molecule type: DNA
 A:Residues: 1-652 <GOL>
 A:Cross-references: GB:U72152; EMBL:M63192; NID:G1763224; PIDN:AAC44766.1; PID:G1763226
 A:Note: The sequence from Fig. 3 is inconsistent with that from Fig. 2 in having 299-Thr
 R:Goldberg, M.B.; Boyko, S.A.; Calderwood, S.B.
 J. Bacteriol. 172, 6863-6870, 1990
 A:Title: Transcriptional regulation by iron of a Vibrio cholerae virulence gene and homo
 A:Reference number: A37834; MUID:91072235; PMID:2174861
 A:Accession: A37834
 A:Molecule type: DNA
 A:Residues: 1-152, 'D' <GO2>
 A:Cross-references: GB:M37773
 C:Genetics:
 A:Gene: irga
 C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolo
 C:Keywords: membrane protein
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:26-652/Product: outer membrane protein irga #status predicted <MAT>
 F:68-214/Domain: tonB-dependent receptor amino-terminal homology <TNN>
 F:367-652/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>
 Alignment Scores:
 Pred. No.: 1,74e-64 Length: 652
 Score: 1052.50 Matches: 258
 Percent Similarity: 51.94% Conservative: 117
 Best Local Similarity: 35.73% Mismatches: 222
 Query Match: 28.20% Indels: 125
 DB: 24 Gaps: 24
 US-10-625-972-4 (1-2091) x S25265 (1-652)
 QY 40 CTCGATTTTTCAGCCAGCAGCATAGCT-----GTCGAGGAGGTGATGATT 87
 Db 17 LeuMetPheSerAlaSerAlaPheAlaGlnAspAlaThrLysThrAspGluThrMetVal 36
 QY 88 GTCTCGCATCCGGCTATGAGAAAAAGTACTGACTAACGCGCCCGCTGTTCTGTGATT 147
 Db 37 ValThrAlaAlaGlyTrpAlaGlnValIleGlnAsnAlaProAlaSerIleSerValIle 56
 QY 148 AGCCAGCGAATTCAGTCCAGTCCAGCTACCCAGATCTCGCGGAGGTCTGAGATCAGTA 207
 Db 57 SerArgGluAspLeuGluSerArgTrpTrpArgAspValThrAspAlaLeuLysSerVal 76
 QY 208 GAGGGTGTGATGTTGAAAGTGTGTACGGGTAAACCGGAGGCTGGAAATCAGCATCCGA 267
 Db 77 ProGlyValThrValThrGlyGlyGlyAspThrThr-----AspIleSerIleArg 93
 QY 268 GGAATCCAGCCAGTTACACGCTGATGATTGATTGTTGCTCGTCAGCGCGGAGCAGT 327

A;Residues: 1-665 <KUR>
 A;Cross-references: GB:AL590842; PIDN:CAC90143.1; PID:g15979363; GSPDB:GN00175
 C;Genetics:

A;Gene: YP01313

C;Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolo

Alignment Scores:

Pred. No.: 8, 42e-63 Length: 665
 Score: 1028.00 Matches: 251
 Percent Similarity: 52.12% Conservative: 130
 Best Local Similarity: 34.34% Mismatches: 224
 Query Match: 27.55% Indels: 126
 DB: 2 Gaps: 20

US-10-625-972-4 (1-2091) x AD0160 (1-665)

QY	22	TCGGTAGTCAATCCCTGCTCGGATTTTCAGCC-----AGCAGCATAGCTCTCGCAGAG	75
DB	19	AlaAlaValileSerSerGlnGlyTyrAlaAlaGluLysThrAsnThrAlaThrProThr	38
QY	76	GATGTGATGATTCTCGGCATCCGCTATGAGAAAAGCTGACTAACCGCAGCCGCGT	135
DB	39	AspThrMetValThrAlaSerGlyPheGlnArgIleGlnAspSerAlaLaser	58
QY	136	GTTCCTGTGATTAGCAGGAGGAATTCAGTCCAGCCAGTACACGATCTGGCGGAGGT	195
DB	59	IleSerValThrArgGluGlnIleGluAsnLysAlaTyrArgAspValThrAspAla	78
QY	196	CTCAGATCAGTAGAGGTGTGGATGTTGAAAGTGGTACGGTAAACCGGAGGCTGGAA	255
DB	79	LeuLysAspValProGlyVal---ValIleThrGlyGlySerThrSer-----Asp	95
QY	256	ATCAGATCCGAGGAATGCCAGCAGTATACCGCTGATAGTGAATGATGTTTCGTGAG	315
DB	96	IleSerIleArgGlyMetAlaAlaLysTyrThrLeuIleValAsnGlyLysArgVal	115
QY	316	GGCGGAGCAGTACGTGCTCCACGCGTTCCTGCCATGATACCGGTTTCATGCC	375
DB	116	AspThrArgSerThrArgProAsnSerAspGlySerGlyIleGluGlnGlyTrpLeuPro	135
QY	376	CCTCTGCGCCCATTCAGGGTATTGAGGTATACAGGGCCGATGCCACACTGTATGCC	435
DB	136	ProLeuAlaAlaIleAspArgIleGluValValArgGlyProMetSerSerLeuTyrGly	155
QY	436	TCTGATCGATGGCGGTGTGTGAATATCATACAGAAAGAAATGCAGACAAATGGCTC	495
DB	156	SerAspAlaMetGlyGlyValIleAsnIleThrArgLysValGlyLysGluTrpHis	175
QY	496	TCTTCCGTCAATCGAGGCTGAATCTGCAGGAAAGCAAAATGGGTAAACAGCAGCAG	555
DB	176	GlyThrValArgAlaAspAlaThrLeuGlnGluAspSerLysSerGlyAspIlePheGln	195
QY	556	TTTAATTTCTGGAGCAGTGGTCCCTTGTGGATGATCTGTACGCTGAGTACGCGGT	615
DB	196	ThrAsnAlaTyrAlaSerGlyProLeuIleAspGlyLeuLeuGlyLysValSerGly	215
QY	616	AGCACACACACAGGT-----CAGGGTTCATCGGTACATCATCTGAGC	657
DB	216	LeuLeuSerHisArgSerGluAspLysIleIleAspGlyTyrAsnGlnGlnArgMetArg	235
QY	658	GATACAGCAGGCGGTATTCCTTATCCCGGAGTCAAGAAATATAATCTTGCTGCA	717
DB	236	AsnGlyThrAlaThrPheThrLeuThrProAspAspAsnAsnGluPheAsp-----	252
QY	718	CGTCTTGATCGNAGCGTCCGAGCAGGATGCTCTGCTTGTATGATGATACACACCGG	777
DB	253	-----PheAspIleGlyHisTyrVal	259
QY	778	CAGCGTTATGATAACCGGATGGCAACTGGGAGTCTGACGGGGGATATGACCGGACC	837
DB	260	GlnAspArgAsnSerThrProGlyArgThrLeuAlaLeuAsnGlyThrAsnSerAspThr	279
QY	838	CTGCGGTATGACGGAAACAAAATTTTCAGTGGCTGCTATGATCATCTTACCTTCGGAACA	897

DB	280	---GlnTyrAspArgAsnAsnTyrAlaValThrHisAsnGlyTyrTyrAspPheGlyAsn	298
QY	898	TGGAATCTGATCTGAAGTGAAGACAGACAGAAATAAGTCTGTAGCTTGTACGCGT	957
DB	299	SerThrSerTyrIleGlnArgAspGluThrArgAsnProSer-----	312
QY	958	GTACTGAAGCGCGACAAATGGGCTTGGCGGTACAGCGCGGAGCTTAAGGAATCGAAC	1017
DB	313	-----ArgGlnMetLysSerValAsp	319
QY	1018	CTTATCTCTGAAT-----TCATTACTCTTACCCCTCTGGGAGAAATCTATCTGTT	1068
DB	320	AsnIlePheAsnThrGlnThrSerPheLeuLeu-----AspAsnHisThrLeu	335
QY	1069	ACGCTGGGCGGAGTTTCAGAGCTCGTCCATGAAAGAC---GGAGTTGCTTCCGACG	1125
DB	336	IleLeuGlyGlyGlnTyrArgTyrGluGluLeuTyrAspLysGlyAsnGlnLeuProSer	355
QY	1126	ACAGGTGAA-----ACTTTCCGCGCAGAAAAGCTGCTCGTATTTCGTAGGATGAGTG	1179
DB	356	AlaSerAspLeuLysLysLeuThrArgTyrSerTyrAlaLeuPheAlaGluAspGluTrp	375
QY	1180	CATCTCAGGATGACCTTGGCTGACTGCGGCGAGCGCTATGAAATCATCATGACCAATTC	1239
DB	376	GlnMetThrAsnAspPheAlaLeuThrGlyGlyIleArgMetAspGlnAspGlnAsnTyr	395
QY	1240	GGGCGACACTTCAGTCCGCTGCATCTGCTGGGATGTGGCAGATGCTCGAGCGCTG	1299
DB	396	GlyThrHisThrProArgLeuTyrGlyValThrHisLeuAlaAspGlnTrpThrLeu	415
QY	1300	AAAGCGGTGTGACCGGATATAAGCACCCAGAAATGGGCGAGCTACATAAAGGAT	1359
DB	416	LysGlyValSerGlyTyrArgSerProAspLeuArgGlnAlaThrAspAspTrp	435
QY	1360	AGTGTGTGTCGGCAGGAGGAAAACAAATCTA-----CTTGTAACCCGAC	1407
DB	436	GlyGlnLeuSerGlyGlyLysGlyLeuProAlaLeuLeuLeuGlyAsnSerAsn	455
QY	1408	CTGAAGCCGGAAGAGCGTCAGTTATGAGCTGGGTGGTATTACGATAACCCCGCGGT	1467
DB	456	LeuLysProGluArgSerIleSerGlnGluIleGlyLeuTrpAspAspGlnGluGly	475
QY	1468	CTGATGCCAATCTCAGGTTTATGACTGCTCTCCAACAGAGTTGTC-----	1518
DB	476	MetAsnAlaSerValThrLeuPheTyrThrAspPheLysAspLysIleThrGluValArg	495
QY	1519	TCTTATTTCCATAATGATAACAC-----	1542
DB	496	AsnCysAspIleThrThrAsnThrThrGlyGlnCysValPheAsnGlyIleAsnTyrLys	515
QY	1543	-----AATAGCTATGTAAACAGCGGAAGGCCGCTTGCACGCTGTGAA-----TTT	1590
DB	516	PheIleSerAspArgIleAsnValAspLysAlaMetThrArgGlyAlaGluAlaThrPhe	535
QY	1591	GCGGCGACATG---CGCTGTGCTCAGAGGATGTCACGCTGTCACTGAATTACACTGG	1647
DB	536	AlaTrpAspIleAsnGlnAlaTrpSer-----LeuAlaThrAsnTyrThrPhe	551
QY	1648	ACCGAAGTGAACACGCTGATGTGTATACAAAGGTGCGCGCTGATATACCCCTGAA	1707
DB	552	ThrGlnSerGluGlnLysSerGlyAlaPheAlaGlyGlnProLeuAsnGlnMetProLys	571
QY	1708	CACATGGTGAATCGAACTGAACTGCAGATCACCGAAGAGGTGTCATCATGGCTGGGT	1767
DB	572	HisMetLeuAsnGlyThrLeuAsnTrpLysThrThrGluAspPheAlaThrTrpIleArg	591
QY	1768	GCCCGTTATCGCGGAAAAACACCGCTTTCACCCAGAAATTTTCGTCATGACGCGCTGA	1827
DB	592	AlaAsnTyrArgGlyLysAlaSerGluTyrLeuAsnArgThrSer-----	606
QY	1828	CAGAAGAAAGTGTATGATGAGAAAGGAGAAATACCTGAAAGCCTGGACGCTGGATGCA	1887


```

Db 451 GluIleArgGlnIleAlaProGlyTyrAlaTyrThrThrGlyGlyGlyCysSerTyr 470
Qy 1390 -----CTACTTGGTAACCCGACCTGAAGCCGGAAGAGAC 1425
Db 471 GlyProSerGlyThrCysGlyValIleIleGlyAspProAsnLeuGluAlaGluYsSer 490
Qy 1426 GTCAATTATGAGCTGGGTGATTACGATAACCCCGCGGTCTGAATGCCAATGTCACA 1485
Db 491 ThrSerTyrGluIleAlaLeuTyrAspAsn---GlyAspIleAlaLeuGlyAlaThr 509
Qy 1486 GGTTTTATGACTGACTTCTCCAAACAAGATT----- 1515
Db 510 TyrPheTyrThrAspPheLysAspLysIleSerAsnAlaLeuValLeuAsnProAspGly 529
Qy 1516 -----GTCTCTTATTCCTATAAT 1533
Db 530 ThrProAlaArgTyrSerGluAspArgAsnTyrArgLeuTyrAsnTyrAsnIleAsp 549
Qy 1534 GATAACACCAATAGCTATGTAACACAGCGAAGCCGCGTGCACGCGTGTGGAATTTGCC 1593
Db 550 Asp-----AlaIleIleGlnGlyValGluLeuThr 559
Qy 1594 GGCACATGTCGCGTGTGG-----TCAGAGGATGTCACGCTGTCATGAATTACACCTGG 1647
Db 560 AlaThr-----TrpTyrAlaThrSerGluLeuThrLeuArgGlyAsnTyrThrTyr 576
Qy 1648 ACCGAGGTGAACAACGTCGTGTGATACAAAGGTGCGCGCTGAGTTATACCCCTGAA 1707
Db 577 ThrHisSerGluGlnIleThrGlyAspTyrGluGlyPheProLeuAlaArgThrProGlu 596
Qy 1708 CATATGTCGAATCGAAATGTAACCTGCGCAGATCACC-----GAAGAGGTGCGATCATGG 1761
Db 597 HisMetAlaAsnLeuArgGlyAspTyr-----ValThrProIleAspGlyLeuGluAlaTyr 615
Qy 1762 CTGGTGGCCGTTATCGCGGAAACA-----CCAGTTTCACCCAGAAATAT 1809
Db 616 ValSerLeuAsnTyrHisGlySerGluIleAsnAlaGlyProArgIleGlyAlaAsnGly 635
Qy 1810 TCCTCACTAGCCCTGTACAGAGAAGTGTATGATGAGAAAGAGAAATACCTGAAGCC 1869
Db 636 ThrProValThr-----IleAsnGlyGlnSerGlyArgLysTyrAspAla 650
Qy 1870 TGACCGGTGTGATGAGCTGTGTCGTGAAGATGACGGATCCCTGACGCTGAATGCT 1929
Db 651 TyrThrThrLeuAspIleGlyAlaLysTyrAlaValAlaGluAsnValAspLeuAsnAla 670
Qy 1930 GCGGTGAATAACCTGTCACCAAGGATTCACGTGACGTGACGCTGTACAGTCCCGGTAAG 1989
Db 671 AlaValTyrAsnValPheAspLysAsp----- 679
Qy 1990 AGTACGCTGTATGCGCGTATCTTCCAGACGGGATCATCAACACAGGATATGTGATA 2049
Db 680 -----ValGlyThrAspAspPheAsnThrValMet 689
Qy 2050 CCTGACGGAATTACTGGATGTCGCTGAACATCATCAGTTC 2088
Db 690 GluGlyArgArgPheThrIleSerMetThrAlaLysPhe 702
RESULT 8
E98247
outer membrane protein irga precursor [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: E98247
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:21608551; PMID:11743194
A:Accession: E98247
A>Status: preliminary
A:Molecule type: DNA

```

```

A:Residues: 1-707 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89503.1; PID:g15159377; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L1858
A:Map position: linear chromosome
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homologs

Alignment Scores:
Pred. No.: 2,17e-60 Length: 707
Score: 993.00 Matches: 242
Percent Similarity: 49.80% Conservative: 133
Best Local Similarity: 32.14% Mismatches: 244
Query Match: 26.61% Indels: 134
DB: 2 Gaps: 23

US-10-625-972-4 (1-2091) x E98247 (1-707)
Qy 19 GCTTCCGTAGTCATTCCTCTCGGATTTTCAGCCAGCAGCATAGCT-----GCTGCA 72
Db 26 AlaAlaMetAlaGlyThrAlaLeuGlyPheAlaValProAlaPheAlaGlnAlaSer 45
Qy 73 GAG-----GATGTGATGATGCTCTCGGCATCCGCTATGAGAAAAAGCTG 117
Db 46 GluGlyAsnThrValLeuGlnGlnIleValThrAlaSerGlyPheGluGlnAsnVal 65
Qy 118 ACTAACCCAGCCGCGCAGTGTCTGTGATTAGCCAGGAGAAATTCAGTCCAGCCAGTAC 177
Db 66 LysAspAlaProAlaSerIleThrValThrArgGluAspLeuGluLysGlySerTyr 85
Qy 178 CACGATCTGCGGAGGCTCTGAGATCAGTAGAGGGTGTGGATGTTGAAAGTGTGCGGT 237
Db 86 ArgAspLeuThrAspAlaLeuArgGluValGlnGlyValSerVal-----ThrGly 102
Qy 238 AAAACCCGCGGCTGGAATCAGCATCCGAGGAATCCAGCCAGTATTCAGCTGATAC 297
Db 103 IleAlaAsnGluLysAspValPheIleArgGlyLeuProGlyAlaTyrThrLeuIleLeu 122
Qy 298 ATTGATCGTGTTCGTACAGCGGAAGCAGTCAGTCACTCCCAACGGTTTTTCGCCATG 357
Db 123 ValAspGlyLysArgGln---SerThrArgAspAlaArgThrAsnGlyAsnSerGlyPhe 141
Qy 358 AATACCCGGTTCATGCCCTCTGCGCCCATTCAGCGTATTCAGGTTATCAGGGGCGG 417
Db 142 GluGlnSerPheValProValSerAlaIleGluArgIleGluValValArgGlyPro 161
Qy 418 ATGTCCACACTGTATGCTCTGATGCGATGGCGGTGTGTGTAATATCATACCAAG 477
Db 162 MetSerSerLeuTyrGlySerAspAlaMetGlyGlyValIleAsnIleIleThrArgLys 181
Qy 478 AATGCAGACAATGGCTCTCTTCGTCATGTCAGGGCTGATCTGCAGGAACACACAA 537
Db 182 ValGlyAspValTrpSerGlySerValThrThrGluGlyThrValGlnGlnHisSerLys 201
Qy 538 TGGGGTAACAGCAGCAGGTTTAATTTCTGGAGCAGTGGTCCCTCTGTGGATGATCTGTC 597
Db 202 PheGlyAsnSerGlyGlnValSerTyrTyrAlaAsnGlyProIleLeuLysAspGlnLeu 221
Qy 598 AGCCTGAGGTACCGGTAGCACACACAGCGTCGAGGGTTTCATCGGTCACTCATGAGC 657
Db 222 GlyLeuGlnLeuTrpGlyArgGlyPheThrArgGlyGluAspArgIle-----LeuAsn 239
Qy 658 GATACAGCAGCAGCGTATTCCTTATCCAGCGATGCACAGAAATATATCTGTGTGCA 717
Db 240 GlyThrThrGly-----AlaLysGluTyrAspPheAsnGly 251
Qy 718 CGTCTTGAAGCGGTGGAGCAGGATGCTCTCTGTTGATATGATGATACCCCGG 777
Db 252 ArgLeuThrPheThrProAsnGluAspHisAspIleTyrLeuGluGlyGlyLysThrArg 271
Qy 778 CACGGTATGATAACCGGATGGCACTGGGAGTCTGACGGGGGATATGACCGGACC 837
Db 272 LeuArg-----ArgAspAlaGluProGlyAspThrLeuAlaAlaThrAspAlaAsn 288

```



```
QY 838 CTGCGCTATAGCGAAAC---AAAATTTCAGCTGGCTATGATCATCATCTTACCTTCGGA 894
Db 289 GlyThrTyrAsnThrAsnThrArgAspHisTrpSerLeuSerHisThrGlyArgTrpGly 308
QY 895 -----ACATGGAATCGTATCTGAAGTGAACGAG 924
Db 309 ProThrThrSerGluPheSerPheGlnGlnGluTrpAlaGluArgThrAsnPheThrArg 328
QY 925 ACAGAAAATAAAGGTCGTGAGCTTGTCAGCAGTGTACTGAAGCGCGCAAAATGGGGCTT 984
Db 329 AsnThrArgThrGlyArg-----Val 335
QY 985 GCCGGTCAGCGCGGAGCTTAAGAAATCGAACCTTATCTTGAATTCATCTACTGCTTACC 1044
Db 336 ThrGluAsnProArgSerProGluLeuArgAsnThrValLeuAspGlyLysPheThrThr 355
QY 1045 CCT-----CTGGGAGATCTCATCTGTTACGGTGGGGCGGAGTTTCAGAGTCGTGCC 1098
Db 356 ProPheGluLeuPheGlyAsnHisThrLeuValThrGlyGlyGlnTyrPheGluAlaArg 375
QY 1099 ATGAAGACGGAGTTGTCTTCCAGCAC-----GGTGAACCTTTCCGGCAGAAAGC 1152
Db 376 LeuThrAspGlnAsnProGlyArgArgThrAspArgAspGluThrPheSerAlaThrGln 395
QY 1153 TGTCGCTATTGCTGAGATGAGTGGCATCTCACGGATGCATTCGCGTACTGCGGGC 1212
Db 396 TrpAlaLeuPheLeuGluAspGluTrpArgIleValAspAsnPheAlaLeuThrGlyGly 415
QY 1213 AGCCCTCATGAACATCATGAGCAATTCGGGGACACTTCAGTCGCGTGCATATCTGTC 1272
Db 416 LeuArgLeuAspAsnHisGluLysTyrGlyAsnHisPheSerProArgLeuTyrGlyVal 435
QY 1273 TGGGATGCGCAGATCCCTGAGCGTGAAGCGGTGTGACCCGGGATATAAGCCACC 1332
Db 436 TrpSerAlaThrGluGluLeuThrIleLysGlyValSerThrGlyPheArgAlaPro 455
QY 1333 AGAATGGGCGACATACATAAAGCGATTAGTGTGTGTCGGGCGAGGAAAAACAAT--- 1389
Db 456 GluIleArgGluIleAlaProGlyTyrAlaTyrThrThrGlyGlyGlyCysSerTyr 475
QY 1390 -----CTACTGTGTAACCCGACCTGAAGCGGAGCAGAGAGC 1425
Db 476 GlyProSerGlyThrCysGlyValIleGlyAspProAsnLeuGluAlaGluLysSer 495
QY 1426 CTCAGTTATGAGCTGGGTGATTACGATACCCCGCGTCTCAATGCCAATGTCACA 1485
Db 496 ThrSerTyrGluIleAlaAlaLeuTrpAspAsn---GlyAspIleAlaLeuGlyAlaThr 514
QY 1486 GGTTTTATGACTCTTCTCCAAACAGATT----- 1515
Db 515 TyrPheTyrThrAspPheLysAspLysIleSerAsnAlaLeuValLeuAsnProAspGly 534
QY 1516 -----GTCTCTATTCATAAT 1533
Db 535 ThrProAlaArgTrpSerGluAspArgAsnTyrArgLeuTrpTyrAsnThrAsnIleAsp 554
QY 1534 GATAACACCAATAGCTATGTAAACAGCGGAGCGCGGTTCACCGGTGGAATTCGCC 1593
Db 555 Asp-----AlaIleIleGlnGlyValGluLeuThr 564
QY 1594 GGCACATTCCCGCTGG-----TCAGAGGATGTCACCTGTCTACTGAATTACCTGG 1647
Db 565 AlaThr-----TrpTyrAlaThrSerGluLeuThrLeuArgGlyAsnTyrThrTyr 581
QY 1648 ACCGGAAGTGACACAGTGTGATACAAAGGTGCGCGCTGAGTTATCCCTGAA 1707
Db 582 ThrHisSerGluGlnLysThrGlyAspTyrGluGlyPheProLeuAlaArgThrProGlu 601
QY 1708 CACATGCTGAATCGGAAACTGACCTGGCAGATCAC-----GAAGAGTGGCATATGG 1761
Db 602 HisMetAlaAsnLeuArgGlyAspTrp---ValThrProIleAspGlyLeuGluAlaTrp 620
QY 1762 CTGGGTGCCGTTATCGCGGGAACA-----CCAGTTTCACCCAGATTAT 1809
```

```
Db 621 ValSerLeuAsnTyrHisGlySerGluIleAsnAlaGlyProArgIleGlyAlaAsnGly 640
QY 1810 TCCTACTACTGACGGCTCTACAGAAAGATGTATGATGAGAAAGAGGAATACCTGAAAGCC 1869
Db 641 ThrProValThr-----ileAsnGlyGlnSerGlyArgLysTyrAspAla 655
QY 1870 TGACCGTGTGTGATCCAGCTGTCTGCTGAAGATCAGCGATCCCTGACGCTGAATGCT 1929
Db 656 TyrThrThrLeuAspIleGlyAlaLysTyrAlaValAlaGluAsnValAspLeuAsnAla 675
QY 1930 GCGGTGAATAACCTGCTCAACAGGATTACAGTACGCTGAGCCTGTACAGTCCGGTAAAG 1989
Db 676 AlaValTyrAsnValPheAspLysAsp----- 684
QY 1990 AGTACGCTGTATCCCGGTGATTTACTTCCAGCGGATCATCAACACAGATATGTATA 2049
Db 685 -----ValGlyThrAspAspPheAsnThrValMet 694
QY 2050 CCTGACGAAATTACTGGATGTGCTGAACTCATCTCAGTTTC 2088
Db 695 GluGlyArgArgPheIleSerMetThrAlaLysPhe 707
RESULT 9
AG0782
colicin I receptor precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AG0782
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov.
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AG0782
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-663 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD02581.1; PID:G16503438; GSPDB:GN00176
C:Genetics:
A:Gene: STY2434
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolo
Alignment Scores:
Pred. No.: 5.59e-50 Length: 663
Score: 841.50 Matches: 235
Percent Similarity: 48.03% Conservative: 107
Best Local Similarity: 33.01% Mismatches: 263
Query Match: 22.55% Indels: 107
DB: 2 Gaps: 21
US-10-625-972-4 (1-2091) x AG0782 (1-663)
QY 61 ATAGCTCTCGAGAGGAT-----GTGATGTTCTCGGATCCGCTATGAGAAA 111
Db 23 ValAlaAlaAlaThrAspAspGlyGluThrMetValValThrAlaSerAlaIleGluGln 42
QY 112 AAGTGTACTAACGACGCGCGGAGTGTCTGATTAGCCAGGAGGAATTCAGTCCAGC 171
Db 43 AsnLeuLysAspAlaProAlaSerIleSerValIleThrGlnGlnAspLeuGlnArgArg 62
QY 172 CAGTACCACCATCTGGCGGAGGCTCTGAGATCATAGAGGGTGTGGATGTT---GAAAGT 228
Db 63 ProValGlnAsnLeuLysAspValLeuLysGluValProGlyValGlnLeuThrAsnGlu 82
QY 229 GGTACGGGTAAACCGGAGGCTGGAATCAGCATCCGAGGATCCAGCCATTACAGC 288
Db 83 GlyAspAsnArgLysGly-----ValSerIleArgGlyLeuAspSerSerTyrThr 99
QY 289 CTGATACTGATTGATGGTGTTCGTGAGGGCGGAGCAGTACGTGACTCCCAACGGTTT 348
```


Db 599 SerAlaThrAlaLysThrProGlyGlyTyrThrIleTrpAsnThrGlyAlaAlaTrpGln 618
 Qy 1903 ATGACGGATCGCTGACGCTGAATGCTGCGGTGAATAACCTGCTCAACAGGATTACAGT 1962
 Db 619 ValThrLysAspValLysLeuArgAlaGlyValLeuAsnLeuGlyAspLysAspLeuSer 638
 Qy 1963 GACGTGACCTGTACATGCGCGGTAGAGTAGCTGTATGCTGCGGTGATTCTCCAGACG 2022
 Db 639 -----ArgAspAspTyr----- 642
 Qy 2023 GGATCATCAACACAGGATATGTATACCTGACGAGAAATACCTGCGTGTGACTAT 2082
 Db 643 -----SerTyrAsnGluAspGlyArgTyrPheMetAlaValAspTyr 657
 Qy 2083 CAGTTC 2088
 Db 658 ArgPhe 659
 RESULT 11
 QREIC
 colicin I receptor precursor - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C>Date: 30-Sep-1990 #sequence revision 05-Dec-1997 #text change 01-Mar-2002
 C:Accession: B64984; A32056; A33868; A28377; C41871; A35408; S24561
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: B64984
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-663 <BLAT>
 A:Cross-references: GB:AE000304; GB:U00096; NID:gl788470; PIDN:AA075216.1; PID:gl788478;
 A:Experimental source: strain K-12, substrain MG1655
 R:Nau, C.D.; Konisky, J.
 J. Bacteriol. 171, 1041-1047, 1989
 A:Title: Evolutionary relationship between the TonB-dependent outer membrane transport protein and the colicin I receptor.
 A:Reference number: A32056; MUID:89123100; PMID:2644220
 A:Accession: A32056
 A:Molecule type: DNA
 A:Residues: 1-146; RCARCSYHHQ, 159, 'NRSEMV', 166, 'YRYRHHGTSR', 182-527, 'N', 529-613,
 A:Note: the authors translated the codon AAT for residue 528 as Ile
 R:Nau, C.D.; Konisky, J.
 J. Bacteriol. 171, 4530, 1989
 A:Reference number: A33868
 A:Contents: corrections
 A:Accession: A33868
 A:Molecule type: DNA
 A:Residues: 144-184; 610-663 <NA2>
 R:Griggs, D.W.; Tharp, B.B.; Konisky, J.
 J. Bacteriol. 169, 5343-5352, 1987
 A:Title: Cloning and promoter identification of the iron-regulated *cir* gene of Escherichia coli.
 A:Reference number: A28377; MUID:88058737; PMID:3316180
 A:Accession: A28377
 A:Molecule type: DNA
 A:Residues: 1-59 <GRI>
 A:Cross-references: GB:M19295; NID:gl45545; PIDN:AAA23581.1; PID:g551794
 A:Note: residues 26-45 were confirmed by protein sequencing
 R:Steffes, C.; Ellis, J.; Wu, J.; Rosen, B.P.
 J. Bacteriol. 174, 3242-3249, 1992
 A:Title: The *lysP* gene encodes the lysine-specific permease.
 A:Reference number: A41871; MUID:92250419; PMID:1315732
 A:Accession: A41871
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-96, 'D', 98-125 <STE>
 A:Cross-references: GB:M89774; NID:9466776; PIDN:AA17054.1; PID:g466779; EMBL:X65029
 R:Griggs, D.W.; Kafka, K.; Nau, C.D.; Konisky, J.
 J. Bacteriol. 172, 3529-3533, 1990
 A:Title: Activation of expression of the Escherichia coli *cir* gene by an iron-independent promoter.
 A:Reference number: A35408; MUID:90264362; PMID:2160948
 A:Accession: A35408

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-21 <GR2>
 C:Comment: This outer membrane receptor for colicins Ia and Ib is regulated by both cell envelope and membrane proteins.
 C:Genetics:
 A:Gene: *cir*; *cirA*; *feuA*
 A:Map position: 43 min
 C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homology
 C:Keywords: iron transport; membrane protein
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:26-624/Product: colicin I receptor #status experimental <VAT>
 F:66-211/Domain: tonB-dependent receptor amino-terminal homology <TN>
 F:365-663/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>
 Alignment Scores:
 Pred. No.: 2,33e-49 Length: 663
 Score: 832.50 Matches: 234
 Percent Similarity: 47.45% Conservative: 120
 Best Local Similarity: 31.37% Mismatches: 237
 Query Match: 22.31% Indels: 155
 DB: 1 Gaps: 25
 US-10-625-972-4 (1-2091) x QREIC (1-663)
 Qy 40 CTCGGATTTTCAGCCAGCAGCATAGCTGCTGCA----- 72
 Db 10 ValGlyLeuCysLeuSerAlaIleSerCysAlaTrpProValLeuAlaValAspAsp 29
 Qy 73 GAGGATGTGATGTTCTCGGCATCCGGCTATGAGAAAAGCTGACTACGCGCGGCC 132
 Db 30 GlyClnThrMetValThrAlaSerSerValGluGlnAsnLeuLysAspAlaProAla 49
 Qy 133 AGTGTCTGTGATAGCAGGAGGAATTCAGTCCAGCCAGTACCGATCTGGCGGAG 192
 Db 50 SerIleSerValIleThrGlnGluAspLeuGlnArgLysProValGlnAsnLeuLysAsp 69
 Qy 193 GCTCTGAGATCAGTAGAGGCTGTGGATGTT---GAAAGTGTGTACGGTAAACCGGAGG 249
 Db 70 ValLeuLysGluValProGlyValGlnLeuThrAsnGluGlyAspAsnArgLysGly--- 88
 Qy 250 CTGAAATCAGCATCCGAGGAATCCGACGATGATACGCTGATCTGATGTTGTTGT 309
 Db 89 -----ValSerIleArgGlyLeuAspSerSerTyrThrLeuIleLeuValAspGlyLys 106
 Qy 310 CGTCAGGCGGAGGAGCAGTGCAGTCCCAACGGTTTCTGCCATGAATACCGGGTTC 369
 Db 107 ArgValAsnSerArgAsnAlaValPheArgHisAsnAspPheAspLeuAsn----- 123
 Qy 370 ATGCCCCCTCTCGCCGCGCATTTAGCGGTATGAGGGGCGCGATGTCACACTG 429
 Db 124 TrpIleProValAspSerIleGluArgIleGluValValArgGlyProMetSerSerLeu 143
 Qy 430 TATGGCTCTCATGCGTGGCGGTGTGGTGAATATCATTACCAAGAAAGATGCAGACAA 489
 Db 144 TyrGlySerAspAlaLeuGlyGlyValValAsnIleIleThrLysLysIleGlyGlnLys 163
 Qy 490 TGGTCTCTCTTCCTCAATGCGCGGCTCAATTCAGGAAAGCAACAAATGGGTAAACAGC 549
 Db 164 TrpSerGlyThrValThrValAspThrThrIleGlnGluHisArgAspArgGlyAspThr 183
 Qy 550 AGCAGTTTAAATTTGAGACAGTGTCCCTTGTGGATGATCTCTGTCAGCTCGAGGTA 609
 Db 184 TyrAsnGlyGlnPhePheThrSerGlyProLeuIleAspGlyValLeuGlyMetLysAla 203
 Qy 610 CGCGGTAGCACACACAGCGT-----CAGGGTTTCGCTCAGTCACTCAGT 654
 Db 204 TyrGlySerLeuAlaLysArgGluLysAspAspProGlnAsnSerThrThrThrAspThr 223
 Qy 655 AGCGATACAGGCGCGGCTATCTCTTATCCACCGAGTCCACAGATTAATTAATCTTGT 714
 Db 224 GlyGluThr----- 226

QY	193	GTCTCTGAGATCAAGTACAGGGTGTGGATGTT---GAAAGTGTGTACGGGTAAACCCGAGGG	249
DB	70	VaileuLysGluValProGlyValGlnLeuThrAsnGluGlyAspAsnArgIysGly---	88
QY	250	CTGGAATCAGATCCGAGGATCCACCCAGTTTACAGCTGATCTACTGATTGATGTGTT	309
DB	89	-----ValSerIleArgGlyLeuAspSerIyrThrLeuIleLeuValaspGlyLys	106
QY	310	CGTCAGGGCGGAGCAGTGAOCTGACTCCCAACCGCTTTTCTGCCATGAATACCGGGTTC	369
DB	107	ArgValAsnSerArgAsnAlaValPheArgHisAsnAspPheAspLeuAsn-----	123
QY	370	ATGCCCCCTCTCGCCGCAATTCAGGGTATTCAGGTATTCAGGGGGCGGATGCCACACGTG	429
DB	124	TpIleProValAspSerIleGluArgIleValValArgGlyProMetSerSerLeu	143
QY	430	TATGCTCTGATCGCATGGCGGTGTGGTGAATATCATTCACAGAAGAAATGCAGACAA	489
DB	144	TyrGlySerAspAlaLeuGlyGlyValValAsnIleIleThrIysIysIleGlyGlnIys	163
QY	490	TGGCTCTCTTCGTCAATGCAGGGCTGNAATCTGCAGGAAGACAAATGGGGTAACAGC	549
DB	164	TpSerGlyThrValThrValaspThrThrIleGlnIleHisArgAspArgGlyAspThr	183
QY	550	AGCCAGTTTAAATTCCTGGAGCAGTGGTCCCTTGTGATGATTCCTGCAGCTGCAGGTA	609
DB	184	TyrAsnGlyGlnPhePheThrSerGlyProLeuIleAspGlyValLeuGlyMetIysAla	203
QY	610	CGGGTAGCACACACACAGCT---CAGGTTTCATCGGTCAATCATCACTG	654
DB	204	TyrGlySerLeuAlaLysArgGluLysAspAspProGlnAsnSerThrThrAspThr	223
QY	655	ACGATACAGCAGGCACCGCTATTCCTTATCCACGAGTCAAGAAATTAATCTTGCT	714
DB	224	GlyGluThr-----	226
QY	715	GCAGCTCTTGATGGAAGCGCTCGAGCAG---GATGTGCTCTGGTTTGTATG	765
DB	227	ProArgIleGluGlyPheSerSerArgAspGlyAsnValGluPheAlaTrp-----	243
QY	766	GATACACCCGCGACGCTTATGATAACCGGATGGGCACTGGGAGTCTGACGGGGGA	825
DB	244	---ThrProAsnGlnAsnHisAspPheThrAlaGlyTyr-----Gly	256
QY	826	TAATGACCGG-----ACCTGGCTATCAGCGAAACAAA	858
DB	257	PheAspArgGlnAspArgAspSerAspSerLeuAspLysAsnArgLeuGluArgGlnAsn	276
QY	859	ATTTCAGTGGCTATGATCATCTTCACCTTCGGAACATGGAAATCGTATCTGAAC	918
DB	277	TyrSerValSerHisAsnGlyArgTrpAspTyrGlyThrSerGluLeuLysTyrTyrGly	296
QY	919	AACGAGACAGAAATAAAGTGTGAGCTGTGACGAGTGTACGAGTGTACTGAACGCGCAGAAATGG	978
DB	297	GluLysValGluAsnLysAsn-----	303
QY	979	GGGCTTGGCGGTACGCGCGGAGCTT---AAGAAATCGAACTTATCTGAATTCATTA	1035
DB	304	-----ProGlyAsnSerSerProIleThrSerGluSerAsnThrValaspGlyLysTyr	321
QY	1036	CTGCTTACCCCTCTGGGGAATCTCATCTCGTACGGTGGGGCGAGTTTTCAGAGCTCG	1095
DB	322	ThrLeuProLeuThrAlaIleAsnGlnPheLeuThrValGlyGlyLysTrpArgHisAsp	341
QY	1096	TCCATGAAGACGGAGTGTCTCTGCCAGACAGAGTGAACCTTCGGCGAGAAAGC---	1152
DB	342	LysLeuSerAspAlaValAsnLeu-----ThrGlyGlyThrSerSerLysThrSerAla	359
QY	1153	-----TGGTCCGTTATTTGTGAGGATGAGTGGCATCTCACGATGCATCTCGCTGACT	1206
DB	360	SerGlnTyrAlaLeuPheValGluAspGluTrpArgIlePheGluProLeuAlaLeuThr	379

QY	1207	CGCGGAGCCGCTATGAACATCATATGACGAANTTCGGGGGACACTTCACTCGCGTGCATAT	1268
Db	380	ThrGlyValArgMetAspAspHisGluThrTyrGlyGluHisTrpSerProArgAlaTyr	399
QY	1267	CTGTCTCGGATGTGGCAGATGCGTCGACGCTGAAGCCGCTGACACCGGATGACACCGGATATAAG	1326
Db	400	LeuValTyrAsnAlaThrAspThrValThrValValGlyGlyTyrPalaThrAlaPheLys	419
QY	1327	GCACCCAGATGGGGCGACTACATAAAGGGATTAGTGTGTCTCGCGGCGAGGAAACA	1386
Db	420	Ala***SerLeuLeuGlnLeuSerProAspTrpThrSerAsnSerCysArgGlyAlaCys	439
QY	1387	AATCTACTTGTTAAACCCGACCTGAAGCCGGAAGAGCGCTCAGTTATGAGGCTGGGGT	1446
Db	440	LysIleValGlySerProAspLeuLysProGluThrSerGluSerTrpGluLeuGlyLeu	459
QY	1447	TATTACGATAACCCC-----GCCGCTCTGAATCCCAATGTCACAGGTTTATG	1494
Db	460	TyrTrpMetGlyGluGluGlyTrpLeuGluGlyValGluSerSerValThrValPheArg	479
QY	1495	ACTGACTTCTCCACAGAGATTGCTCTTATTCATAAATGATAACCAAT-----	1545
Db	480	AsnAspValLysAspArgIle---SerIleSerArgThrSerAspValAsnAlaAlaPro	498
QY	1546	-----ACGTATGTAAAC	1557
Db	499	GlyTyrGlnAsnPheValGlyPheThrAlaAsnGlyValProValPheSerTyrTyrAsn	518
QY	1558	AGCGGAAGCCCGGTTCACGGTGTGGAAATTTGCGGCACATTCGCGTGTGGTCAGAG	1617
Db	519	ValAsnLysAlaArgIleGlnGlyValGluThrGluLeuLysIlePro---PheAsnAsp	537
QY	1618	GATGTCAGCTGTCTACTGAATTCACACTGGACCCGAGTGACAAAGTGAT-----	1668
Db	538	GluTrpLysLeuSerLeuAsnTyrThrTyr-----AsnAspGlyArgAspValSerAsn	555
QY	1669	GGTGATAACAAAGTGGCGCGCTGAGTTATACCCCTGAACACATGTTGTAATGCGAACTG	1728
Db	556	GlyGluAsnLys-----ProLeuSerAspLeuProPheHisThrAlaAsnGlyThrLeu	573
QY	1729	AATCGGCAG-----ATCACCGAAGAGTGGCATCATGGCTGGGTCCCGCTTATCCGGG	1782
Db	574	AspTrpLysProLeuAlaLeuGluAspTrpSerPheTyrValSerGlyHisTyrThrGly	593
QY	1783	AAAACACACGCTTTCACCCAGAAATTATTGTCACCTGAGCGCTGTACAGAAAGATGTAT	1842
Db	594	-----GlnLysArgAlaAsp	598
QY	1843	GATGAGAAGGAGAAATACCTGAAGCGTGGATGCGTGCAGTCTGTCGTGAAG	1902
Db	599	SerAlaThrAlaLysThrProGlyTyrThrIleTrpAsnThrGlyAlaAlaIleTrpGln	618
QY	1903	ATCACGATGCCCTGACGCTGAATGCTGGGTGAATAACCTGCTCAACAGGATTACAGT	1962
Db	619	ValThrLysAspValLysLeuArgAlaGlyValLeuAsnLeuGlyAspLysAspLeuSer	638
QY	1963	GAGCTGAGCCTGTACAGTCCCGGTAAAGATACGCTGTATGCCGGTGATTACTTCCAGACG	2022
Db	639	-----ArgAspAspTyr-----	642
QY	2023	GGATCATCAACACAGGATATGTGATACCTGACGCGAAATTACTGGATGTCGCTGAACTAT	2082
Db	643	-----SerTyrAsnGluAspGlyArgTyrPheMetAlaValAspTyr	657
QY	2083	CAGTTC	2088
Db	658	ArgPhe	659

C
 T
 E
 H
 C
 D
 D

F83405
 probable TonB-dependent receptor PA1922 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:date: 15-Sep-2000 #sequence revision 15-Sep-2000 #text change 31-Dec-2000
 RESULTS 13

QY 1729 AACTGGCAGATCACCGAAGAGTGGCATCATGGCTGGTCCCGTTATCCGCGGAAACA 1788
 Db : : : : :
 QY 567 ArgTrpSerProAnValAlpGheSerAlaGlnLeuArgThrGluTyrValGlySerGln 586
 Db : : : : :
 QY 1789 CCAGTTTCCACCCAGATTATTCGTCACTAGCGCTGTACAGAAAGTGTATGATGAG 1848
 Db : : : : :
 QY 587 ValAlaTyrSerSerAnVal----- 593
 QY 1849 AAAGGAGATACCTGAAAGCTCGACGGTGGTGGATGACAGGTCTGTGCGGAGATGACG 1908
 Db : : : : :
 QY 594 --GlyTyrAlaLeuProAlaTyrSerLeuTrpHisLeuGluLeuSerGlnLysLeuSer 612
 QY 1909 GATGCCCTGAGCTGAATGCTGGGTGTAATAACTGCTCAACAGGATTACAGTGCAC 1965
 Db : : : : :
 QY 613 GluAsnLeuThrLeuArgGlyIleGluAsnLeuGlyAspGlnArgLeuAlaAsp 631
 RESULT 14
 G90706
 hypothetical protein ECs0623 [imported] - Escherichia coli (strain O157:H7, substrain R)
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision: 18-Jul-2001 #text_change: 03-Aug-2001
 C:Accession: G90706
 R:Hayaashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C. G.
 gasawashi, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
 A:Reference number: A99629; PMID:21156231; PMID:11258796
 A:Accession: G90706
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-746 <HAY>
 A:Cross-references: GB:BA000007; PIDM:BAE34046.1; PID:G13360081; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RMD 0509952
 C:Genetics:
 C:Gene: ECs0623
 C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolo
 Alignment Scores:
 Pred. No.: 2,33e-35 Length: 746
 Score: 629.00 Matches: 210
 Percent Similarity: 45.25% Conservative: 133
 Best Local Similarity: 27.70% Mismatches: 285
 Query Match: 16.85% Indels: 130
 DB: 2 Gaps: 29
 US-10-625-972-4 (1-2091) x G90706 (1-746)
 QY 73 GAGGATGATGATGATTCCTCGCATCCGCGTATGAGAAAGCTGACTAACGACGCCGCC 132
 Db : : : : :
 QY 33 AspAspThrIleValValThrAlaAla-----GluGlnAsnLeu-----GlnAlaPro 48
 QY 133 AGTGTTCCTGTGATTACCCAGGAGGAATGTCAGTCCAGCCAGTAC---CACGATCTGGC 189
 Db : : : : :
 QY 49 GlyValSerThrIleThrAlaAspGluIleArgLysAsnProValAlaArgAspValSer 68
 QY 190 GAGGCTCTGAGTACGATGAGGCTGTGGATGT-----GAAAGTGGTACGGTAAAC 243
 Db : : : : :
 QY 69 GluIleArgThrMetProGlyValAsnLeuThrGlyAsnSerThrSerGlyGlnArg 88
 QY 244 GGA---GGCTGGAATCAGTACGAGGCTGTGGATGT-----GAAAGTGGTACGGTAAAC 300
 Db : : : : :
 QY 89 GlyAsnAsnArgGlnIleAspIleArgGlyMetGlyProGluAsnThrLeuIleLeuIle 108
 QY 301 GATGGT-----GTTCTGACGGCGGGAAGCAGTGCAGCTGACT 336
 Db : : : : :
 QY 109 AspGlyLysProValSerSerArgAsnSerValArgGlnGlyTrpArgGlyGluArgAsp 128
 QY 337 CCCAAGCTTTTTCGCCATGATACCGGGTTCATGCCCTCTGCCGCCATTCAGGCT 396
 Db : : : : :
 QY 129 ThrArgGly-----AspThrSerTrpValProPro---GluMetIleGluArg 143
 QY 397 ATTGAGGTTATCAGGGGCGCGATGTCCACACTGTATGGCTCTGTATGCGATGGCGGTGTG 456

144 IleGluValLeuArgGlyProAlaAlaAlaArgTyrGlyAsnGlyAlaAlaGlyGlyVal 163
 QY : : : : :
 QY 457 GTGAATATCATTACCAAGAAAGTGCAGACAAATGGCTCTCTCCGTCATGACGGGCTG 516
 Db : : : : :
 QY 154 ValAsnIleIleThrLysGlySerGlyIleThrLysGlySerTrpAspAlaTyrPhe 183
 QY : : : : :
 QY 517 AATCTGCAGGAAACAAATGGGTAACAGACGCCAGTTTAAATTTCTCGAGCAGTGGT 576
 Db : : : : :
 QY 184 AsnAlaProGluHisLysGluGluGlyAlaThrLysArgThrAsnPheSerLeuThrGly 203
 QY : : : : :
 QY 577 CCCCTGTGGATGATTCTGTGACGCTGCAGGTACCGGTAGCACACACACAGCCTCAGGT 636
 Db : : : : :
 QY 204 ProLeu---GlyAspGluPheSerPheArgLeuTyrGlyAsnLeuAspLysThrGlnAla 222
 QY : : : : :
 QY 637 TCATCG-----GTCACTCACTGACGCGATACAGCAGGACGACGCGCT 675
 Db : : : : :
 QY 223 AspAlaTrpAspIleAsnGlnGlyHisGlnSerAlaAlaAlaGlyThrTyrAlaThrThr 242
 QY : : : : :
 QY 676 ATTCTTATCCACCGAGTGCACAG---AATATATATCTTGTGTCACCTCTGATGGAAG 732
 Db : : : : :
 QY 243 LeuProAlaGlyArgGluGlyValIleAsnLysAspIleAsnGlyValValArgTTPAsp 262
 QY : : : : :
 QY 733 CGCTCGGAGCAGGATGTCTCTGCTGTTTGTATGATACCATCCCGCGAG----- 780
 Db : : : : :
 QY 263 PheAlaProLeuGlnSerLeuGluAlaGlyTyrSerArgGlnGlyAsnLeuTyr 282
 QY : : : : :
 QY 781 -----CGTTATGATAACCGGATGGCAATCTGAGTACCGGGGATAT 828
 Db : : : : :
 QY 283 AlaGlyAspThrGlnAsnThrAsnSerAspAlaTyrThrArgSerLysTyrGly----- 300
 QY : : : : :
 QY 829 GACCGGACCTCGCTATGACGCGAACAATAATTCA-----GCTGCTATGAT 876
 Db : : : : :
 QY 301 AspGluThrAsnArgLeuTyrArgGlnAsnTyrSerLeuThrTrpAsnGlyGlyTTPAsp 320
 QY : : : : :
 QY 877 CATACTTTCACCTTCGGAACATCGTATCTGAACTGGAACGAGACAGACAGAAATAAA 936
 Db : : : : :
 QY 321 AsnGlyValThrThrSerAsnTrpValGlnTyr-----GluHisThrArgAsn--- 336
 QY : : : : :
 QY 937 GGTCTGAGCTTGTACGAGTGTACTGACGCGGACAAATGGGGGCTTCCGGT----- 990
 Db : : : : :
 QY 337 -----SerArgIleProGluGlyLeuAlaGlyGlyThr 347
 QY : : : : :
 QY 991 -----CAGCGCGGAGCTTAAGGAATCGAACCTT----- 1020
 Db : : : : :
 QY 348 GluGlyLysPheAsnGluLysAlaThrGlnAspPheValAspAsnAspLeuAspVal 367
 QY : : : : :
 QY 1021 ATCTGTAATTCATTCTGCTTACCCCTCTGGGA-----GAATCTCATCTGGTTCGGTG 1074
 Db : : : : :
 QY 368 MetLeuHisSerGluValAsnLeuProIleAspPheLeuValAsnGlnThrLeuThrLeu 387
 QY : : : : :
 QY 1075 GGGGGGAGTTTCAGAGCTCGTCCATGAAA----- 1104
 Db : : : : :
 QY 388 GlyThrGluTrpAsnGlnArgMetLysAspLeuSerSerAsnThrGlnAlaLeuThr 407
 QY : : : : :
 QY 1105 -----GACGAGTGTCTTCCACGACAGGTGAACCTTTC 1140
 Db : : : : :
 QY 408 GlyThrAsnThrGlyGlyAlaIleAspGlyValSerAlaThrAspArgSerProTyrSer 427
 QY : : : : :
 QY 1141 CGGCAGAAAGCTGGTGGTATTGCTGAGGATGAGTGCATCTCAGGATCCACCTTCGCG 1200
 Db : : : : :
 QY 428 LysAlaGluIlePheSerLeuPheAlaGluAsnAsnMetGluLeuThrAspSerThrIle 447
 QY : : : : :
 QY 1201 CTGACTCGCGGACCGCTCATGAACATCATGACAAATTCGGGGACACTTCAGTCCGCT 1260
 Db : : : : :
 QY 448 ValThrProGlyLeuArgPheAspHisHisSerIleValGlyAsnAsnTrpSerProAla 467
 QY : : : : :
 QY 1261 GCATATCTGCTGGGATGTGGCAGATGCTGAGCGCTGAAAGCGGTGTGACCGGGA 1320
 Db : : : : :
 QY 468 LeuAsnIleSerGlnGlyLeuGlyAspPheThrLeuLysMetGlyIleAlaAlaArgAla 487
 QY : : : : :
 QY 1321 TATAAGGCAACCCAGAAATGGGCGAGCTACAT-----AAAGGGATT 1359
 Db : : : : :

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: October 13, 2004, 11:54:23 ; Search time 136 Seconds

(without alignments)
9995.325 Million cell updates/sec

Title: US-10-625-972-4

Perfect score: 3732

Sequence: 1 atgggaataaacactctggc.....cgctgaactatcagttctga 2091

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1351062 seqs, 321799191 residues

Total number of hits satisfying chosen parameters: 2702124

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10625972/runat_12102004_130108_26931/app_query.fasta_1.2247
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10625972 -CGN_1_180 -grunat_12102004_130108_26931
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/prodata/1/pubpaa/FCI_NEW_PUB.pcp.*
3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pcp.*
17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pcp.*
18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
------------	-------	-------------	-----------	----	-------------

1	972	26.0	682	14	US-10-238-075-1077	Sequence 1077, Ap
2	428.5	11.5	687	12	US-10-282-122A-48522	Sequence 48522, A
3	399	10.7	684	12	US-10-282-122A-48915	Sequence 48915, A
4	388	10.4	660	14	US-10-238-075-732	Sequence 732, App
5	371	9.9	698	12	US-10-282-122A-77625	Sequence 77625, A
6	355.5	9.5	713	14	US-10-238-075-363	Sequence 263, App
7	347.5	9.3	654	14	US-10-238-075-527	Sequence 527, App
8	321	8.6	713	12	US-10-282-122A-77578	Sequence 77578, A
9	321	8.6	775	12	US-10-282-122A-48409	Sequence 48409, A
10	302.5	8.1	774	12	US-10-282-122A-56053	Sequence 56053, A
11	302	8.1	729	12	US-10-282-122A-75216	Sequence 75216, A
12	301	8.1	763	12	US-10-282-122A-48639	Sequence 48639, A
13	300.5	8.1	774	12	US-10-282-122A-43232	Sequence 43232, A
14	298	7.9	722	12	US-10-282-122A-55554	Sequence 55554, A
15	293.5	7.9	755	12	US-10-282-122A-69342	Sequence 69342, A
16	293	7.9	718	12	US-10-282-122A-4987	Sequence 4987, A
17	290.5	7.8	696	12	US-10-282-122A-76306	Sequence 76306, A
18	283.5	7.6	769	12	US-10-282-122A-48576	Sequence 48576, A
19	283.5	7.6	784	12	US-10-282-122A-66570	Sequence 66570, A
20	283.5	7.6	825	12	US-10-282-122A-51169	Sequence 51169, A
21	281	7.5	699	12	US-10-282-122A-77914	Sequence 77914, A
22	278	7.4	747	12	US-10-282-122A-43354	Sequence 43354, A
23	272	7.3	820	12	US-10-282-122A-66383	Sequence 66383, A
24	270.5	7.2	735	12	US-10-282-122A-60017	Sequence 60017, A
25	269	7.2	687	14	US-10-181-319-2	Sequence 2, Appl
26	267.5	7.2	813	9	US-09-815-242-5085	Sequence 5085, Ap
27	267.5	7.2	813	12	US-10-282-122A-43497	Sequence 43497, A
28	267.5	7.1	704	12	US-10-282-122A-67879	Sequence 67879, A
29	264	7.1	743	12	US-10-282-122A-69539	Sequence 69539, A
30	263.5	7.1	252	11	US-09-764-868-710	Sequence 710, App
31	263	7.0	274	9	US-09-764-868-707	Sequence 707, App
32	263	7.0	774	12	US-10-282-122A-68080	Sequence 68080, A
33	250.5	6.7	926	15	US-10-418-861B-18	Sequence 18, Appl
34	247.5	6.6	767	12	US-10-335-977-4925	Sequence 4925, Ap
35	247.5	6.6	812	12	US-10-282-122A-68347	Sequence 68347, A
36	246	6.6	669	12	US-09-809-665A-105	Sequence 105, App
37	243	6.5	721	12	US-10-282-122A-69343	Sequence 69343, A
38	242.5	6.5	727	12	US-10-282-122A-69051	Sequence 69051, A
39	241.5	6.5	727	12	US-10-282-122A-49650	Sequence 49650, A
40	240.5	6.4	228	9	US-09-764-868-1123	Sequence 1123, Ap
41	240.5	6.4	228	11	US-09-764-875-1022	Sequence 1022, Ap
42	239	6.4	793	12	US-10-335-977-5080	Sequence 5080, Ap
43	239	6.4	793	12	US-10-335-977-5081	Sequence 5081, Ap
44	236	6.4	970	14	US-10-331-061-5	Sequence 5, Appl
45	236	6.3	791	12	US-10-282-122A-69538	Sequence 69538, A

ALIGNMENTS

RESULT 1

US-10-238-075-1077

; Sequence 1077, Application US/10238075

; Publication No. US20030148324A1

; GENERAL INFORMATION:

; APPLICANT: I.N.S.E.R.M.

; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolat

; FILE REFERENCE: BLANDINE

; CURRENT APPLICATION NUMBER: US/10/238.075

; PRIOR FILING DATE: 2002-09-10

; PRIOR APPLICATION NUMBER: 0003145

; PRIOR FILING DATE: 2000-03-10

; NUMBER OF SEQ ID NOS: 1576

; SOFTWARE: Patent in version 3.1

; SEQ ID NO 1077

; LENGTH: 682

; TYPE: PRT

; ORGANISM: Escherichia coli

US-10-238-075-1077

Result

No.

Score

Query Match

Length DB

ID

Description

Alignment Scores:

Pred. No.:

Score:

2.66e-78

972.00

Length:

Matches:

682

244

D	b		306	GlnGluLysValIleArgGluValLysSerGlyLysLysAspLysTyrrAsnHisTrpAsp	329
Q	y		982	CTTCCGGTCAGCGCGGACCTTAAGGAATCGAACCTTATCCTCGAATTCAATTA	1041
D	b		326	LeuAsnTyrGluSerArgLysProGluIleThrAsnThrllelileAspAlalaLysValThr	345
Q	y		1042	ACCCTCTGGAGAAATCTCATCTGGTTACGGTGGGGCGGAGTTTCAGAGCTCGTCGCATG	1101
D	b		346	AlaPheLeuProGlu---AsnValLeuThrIleGlyGlyGlnPheGlnHisAlaGluLeu	364
Q	y		1102	AARGACGGAGTTGTCTTGCCAGCACAGGTCAA-----ACT	1137
D	b		365	ArgAspAsp-----SerAlaThrGlyLysLysThrThrGluThrGlnSerValSer	381
Q	y		1138	TTCGGGAGAAAAGCTGCTCGTATTTCGTGAGGATGAGTGGCATCTCACGGATGCACCTT	1197
D	b		382	IleLysGlnLys----AlaValPheIleGluAsnGluTyrAlaAlaThrAspSerLeu	399
Q	y		1198	GGCGTAGCTCGGGCAGCCGCTATGAACATCATGAGCAATTCGGGGACACTTCAGTCCG	1257
D	b		400	AlaLeuThrGlyGlyLeuArgLeuAspAsnHisGluIleTyrGlySerTyrTrpAsnPro	419
Q	y		1258	CGTGCATATCTGGTCTGGGATGTGGCAGATGCCGTGAGCGGTGAAGACGGGTGCACCA	1317
D	b		420	ArgLeuTyrAlaValTyrAsnLeuThrAspAsnLeuThrLeuLysGlyGlyIleAlaLys	439
Q	y		1318	GGATATAAGGACCCAGAATGGGCGACGCTACATAAAGGGAATTAGTGTGTGTCGGCG	1377
D	b		440	AlaPheArgAlaProSerIleArgGluValSerProGlyPheGlyThrLeuThrGlnGly	459
Q	y		1378	GGAAAAACAATCTACTGTGTAAACCCGACCTGAAGCCGGAAGAGACGTCAGTTATGAG	1437
D	b		460	GlyAlaSerIleMetTyrGlyYasnArgAspLeuLysProGluThrSerValThrGluGlu	479
Q	y		1438	GCTGGGTGTATTACGATAACCCCGCGGTCTGAATGCCAATGTCACAGTTTATGACT	1497
D	b		480	IleGlyIleIeyrSerAsnAspSerGlyPheSerAlaSerAlaThrLeuPheAsnThr	499
Q	y		1498	GACTTCTCCACCAAGATTGCTCTTATTCATCAATAATGAT-----	1536
D	b		500	AspPheLysAsnLysLeuThrSerTyrAspIleGlyThrLysAspProValThrGlyLeu	519
Q	y		1537	AACACCAATAGCTATGTAAACACGCGGAAGCCCGGTGTCACCGTGTGGAAATTCGCG	1596
D	b		520	AsnThrPheIleTyrAspAsnValGlyGluAlaAsnIleArgGlyValGluLeuAlaThr	539
Q	y		1597	ACATTCCGCTGTGGTCAGNGAGTGTACCGCTGTCACTGNAATTACACCTCGACCCG	1656
D	b		540	GlnIleProValTyr---AspLysTrpHisValSerAlaAsnTyrThrPheThrAspSer	558
Q	y		1657	GAACCAACGTGATGGTGATAAC-----AAAGTGC CGCGCTGAGT	1695
D	b		559	ArgArgLysSerAspAspGluSerLeuAsnGlyLysSerLeuLysGlyGluProLeuGlu	578
Q	y		1696	TATACCGCTGAACACATGTTGAATGCGAACTGAATCGCAGATCATCCGGAAGAGTGG	1755
D	b		579	ArgThrProArgHisAlaAlaAsnAlalaLysLeuGluTrpAspTyrThrGlnAspIle	598
Q	y		1756	TCATGGCTGGTGCCCGTTATCGCGG-----	1782
D	b		599	PheTyrSerSerLeuAsnTyrThrGlyLysGlnIleTrpAlaAlaGlnArgAsnGlyAla	618
Q	y		1783	AAAACACACCGTTTACCACGAATTTATCGTCACTGAGCGCTGTACAGAAGAAGTGTAT	1842
D	b		619	LysValProArgValArgAsnGlyPheThrSerMet-----	630
Q	y		1843	GATGAGAAAGGAATACCTGAAAGCCTGACGGTGTGTGATGCAGCTGTCTCGTGGAG	1902
D	b		631	-----AspIleGlyLeuAsnTyrGln	637
Q	y		1903	ATGACGGATGCCCTGACGTGAATGCTCGGTGAATAACCTGCTCAACAGGATTACAGT	1962
D	b		638	IleLeuProAspThrLeuIleAsnPheAlaValLeuAsnValThrAspArgLysSerGlu	657

Db 309 SerGluAspTyrThrTyrGlySerIleAspTyr-----ThrIleSer 323
|||:||||| ||| ||| :|||
Qy 934 AAAGGTGGTGGTGTGACGAGTGA-----CTGAAGCGCGCAAAATGG 978
|||:||||| ||| ||| :|||
Db 324 AspLysSerSerLeuValAlaSerValHisThrAspHisTyrAsnArgTyrAspLysTyr 343
|||:||||| ||| ||| :|||
Qy 979 GGGCTT---CCCGGTGACCGCGGAGCTTAAGAAATCGAACCTTATCTCGTAATTCATTA 1035
|||:||||| ||| ||| :|||
Db 344 GluLeuLysSerGlyArgLeuGluTyrLysAsnAsnIleIleGlnProArgIleVal 363
|||:||||| ||| ||| :|||
Qy 1036 CTGCTTACCCCTCTGGGAGAAATCTCATCTGGTTACGGTGGGGCGGAGTTTCAGAGCTCG 1095
|||:||||| ||| ||| :|||
Db 364 TyrSerThrThrAlaLeuAspLysGlnThrIleThrGlyGlyLeuGluTyrArgGlu 383
|||:||||| ||| ||| :|||
Qy 1096 TCCATGAAAGCAGGAGTGTCTTGGCAGCAGAGGTGAACCTTTCGGCGCAGAAAAGCTGG 1155
|||:||||| ||| ||| :|||
Db 384 SerLeuPheSerAspLysPhe-----GluThrGlyValLysGluAsnLysSerGlnTyr 401
|||:||||| ||| ||| :|||
Qy 1156 -----TCGGTATTTCGAGGATGAGTGCATCTCACGGATGCACCTGGCTGACTGCG 1209
|||:||||| ||| ||| :|||
Db 402 TyrAlaThrAlaPheLeuGlnAspAspTrpSerIleAsnLysGlnPheSerValIleAla 421
|||:||||| ||| ||| :|||
Qy 1210 GCGAGCGCTATGAACATCATGAGCAATTCGGGGGACATTCAGTCCGGTGCATATCTG 1269
|||:||||| ||| ||| :|||
Db 422 GlyLeuArgCysAspTyrHisGluLysTyrGlyThrAsnLeuThrProLysAlaSerVal 441
|||:||||| ||| ||| :|||
Qy 1270 GTCTGGGATGGCAGATGCTGGAGCTGCAAGGGCGGTGTGACACCGGGATATAAGCA 1329
|||:||||| ||| ||| :|||
Db 442 MetTyrLysIlePhe---ProPheThrValArgPheAsnTyrAlaArgGlyTyrArgSer 460
|||:||||| ||| ||| :|||
Qy 1330 CCAGAAATGGCGGAGCTACATAAGGATTAGTGTGTCTGGCGGAGGCAAAAACAAT 1389
|||:||||| ||| ||| :|||
Db 461 ProSerIleLysGluLeuTyrMetAsnTyrAspHisLeu-----GlyMetPheTrp 477
|||:||||| ||| ||| :|||
Qy 1390 CTACTTGGTAACCCGACCTGAAGCGGAGAGAGC-----GTCAATTATGAGCTGGG 1443
|||:||||| ||| ||| :|||
Db 478 IleTyrGlyAsnSerLysLeuLysProGluThrAsnAsnTyrIleSerLeu---SerGly 496
|||:||||| ||| ||| :|||
Qy 1444 GTGTATTACCATACCCCGCGCTCGAATGCCAATGTACACAGTTTATGACTGACTTC 1503
|||:||||| ||| ||| :|||
Db 497 GluTyrValAsnSerTrpIleAsnIleAsnAlaAsnVal-----TyrSerAsnTrpPhe 514
|||:||||| ||| ||| :|||
Qy 1504 TCCAAAGAGATTCTCTTATTCATTAATGATATAACACCAAT---AGCTATTAAACAGC 1560
|||:||||| ||| ||| :|||
Db 515 ArgAsnLysIleGluGlyMetTrpSerAsnAspGlnThrGluLeuHisTyrIleAsnIle 534
|||:||||| ||| ||| :|||
Qy 1561 GGAAGCGCGGTTCACGCTGTGGAATTCGCGGCACATTCGCGTGTGGTCAGAG--- 1617
|||:||||| ||| ||| :|||
Db 535 GlyLysSerArgLeuAlaGlyValGluThrMetCysLysIleGlnIleAsnArgHisIle 554
|||:||||| ||| ||| :|||
Qy 1618 GATGTCACGCTGTCAGTAATTACACCTGGACCCGGAAGTGAACACGTTGATGATTAAC 1677
|||:||||| ||| ||| :|||
Db 555 AsnValHisGlyAlaTyrAsnTyrLeuTyrThr-----SerLysAspAlaAsp--- 570
|||:||||| ||| ||| :|||
Qy 1678 AAAGTGGCCGCTGATGATTATACCCCTGAACATGGTGAATGCCAACTGAACCTGGCAG 1737
|||:||||| ||| ||| :|||
Db 571 ---GlyValArgLeuSerSerSerProHisSerGlyAsnIleArg----- 585
|||:||||| ||| ||| :|||
Qy 1738 ATCACCAGAGAGGTGCATCATGCTGGTGGTGGTTCGCGGGAACACACCACTGTTTC 1797
|||:||||| ||| ||| :|||
Db 586 -----AlaGluTyrAsnThrArgIleProArgTyr 595
|||:||||| ||| ||| :|||
Qy 1798 ACCAG-----AATTATTCGTCATGACGCGCTGTACAGAGAAA-----GTGTATGAT 1845
|||:||||| ||| ||| :|||
Db 596 AlaThrValValAsnLeuSerGlyAsnIleMetGlyLysLysLysPheAspValLeuAsp 615
|||:||||| ||| ||| :|||
Qy 1846 GAG-----AAAGGAGAAATACCTGAAGAGCTGG-----ACGGTGGTGGATCAGGT 1890
|||:||||| ||| ||| :|||
Db 616 GluLeuGluIleAspGlyLysLysValGluAlaTyrTyrGlnAlaLysValAsnProTyr 635
|||:||||| ||| ||| :|||
Qy 1891 CTGCTGTGGAAGATGAC-----GATGCCCTGACGCTGAATCTCGG 1932
|||:||||| ||| ||| :|||

Db 636 CysLeuTrpAspLeuThrValSerGlnTyrIleMetGlnAsnLeuArgIleThrAlaGly 655
Qy 1933 GTGAATAACCTGCTCAACAAGGATTACAGTGACGTGACGCTGTACAGTCCGGTAAAGAGT 1992
|||:||||| ||| ||| :|||
Db 656 IleThrAsnLeuPheAsp-----TyrThrSerAspArgVal 667
|||:||||| ||| ||| :|||
Qy 1993 ACGCTGTATGCGGTGATTACTTCCAGACGGGATCATCAACACAGGATATGTGATACCT 2052
|||:||||| ||| ||| :|||
Db 668 Thr-----PheAsnThrSerThrSer-----Pro 675
|||:||||| ||| ||| :|||
Qy 2053 GAGCGAAATTACTGGATGTCGCTGAACAT 2082
|||:||||| ||| ||| :|||
Db 676 GlyArgAsnTyrPheIleAlaCysAsnTyr 685
|||:||||| ||| ||| :|||
RESULT 3
US-10-282-122A-48915
; Sequence 48915, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48915
; LENGTH: 684
; TYPE: PRT
; ORGANISM: Bacteroides fragilis
US-10-282-122A-48915
Alignment Scores:
Pred. No.: 1,28e-26 Length: 684
Score: 399.00 Matches: 164
Percent Similarity: 42.57% Conservative: 131
Best Local Similarity: 23.67% Mismatches: 264
Query Match: 10.69% Indels: 134
Gaps: 34
US-10-625-972-4 (1-2091) x US-10-282-122A-48915 (1-684)

APPLICANT: I. N. S. E. R. M.
 TITLE OF INVENTION: Polynucleotides which are of nature B2/D4-A- and which are isolated
 FILE REFERENCE: E. coli, and biological uses of these polynucleotides and of their

CURRENT APPLICATION NUMBER: US/10/238,075
 CURRENT FILING DATE: 2002-09-10

PRIOR APPLICATION NUMBER: 0003145
 PRIOR FILING DATE: 2000-03-10

NUMBER OF SEQ ID NOS: 1576
 SOFTWARE: Patent in version 3.1

SEQ ID NO 732

LENGTH: 660

TYPE: PRT

ORGANISM: Escherichia coli

US-10-238-075-732

Alignment Scores:
 Pred. No.: 1,24e-25 Length: 660
 Score: 388.00 Matches: 183
 Percent Similarity: 40.17% Conservative: 107
 Best Local Similarity: 25.35% Mismatches: 272
 Query Match: 10.40% Indels: 160
 DB: 14 Gaps: 37

US-10-625-972-4 (1-2091) x US-10-238-075-732 (1-660)

QY	1	ATCGAATAACCACTCTGGCT-----TCCGTAGTCATCCCTGCTCCGGATTTC	51
DB	9	LeuArgLeuSerLeuLeuAlaValSerAlaThrLeuProThrPheAlaPhe---	27
QY	52	GCACGACGATAGCTGTCGAGAGGATGATGATTCCTCGGCATCCCGCTATGAGAA	111
DB	28	-----AlaThrGluThrMetThrValThrAlaThrGlyAsnAlaArg	41
QY	112	AAGTGACTAACCGCGCCAGTGTCTGTGATTAGCCAGGAGGAAATGAGTCCAGC	171
DB	42	SerSerPheGluAlaProMetValSerValIleAspThrSerAlaProGluAsnGln	61
QY	172	CAGTACCAATACCGGAGGCTCTGAGATCAGTAGAGGCTGTGGATGTTGAAAGTGT	231
DB	62	ThrAlaThrSerAlaThrAspLeuLeuArgHisValProGlyIleThrLeuAsp--	80
QY	232	ACGGTAAACCGAGGCTGGAATCAGCATCCGAGGATCCAGCCAGTACACGGTG	291
DB	81	ThrGlyArgThrAsnGlyGlnAspValAsnMetArgGlyTyrAspHisArgGlyVal	100
QY	292	ATACTGATTGATGTTGCTCAGCGCGGAAGCAGTACGTCAGTCCCAACCGTTTCT	351
DB	101	ValLeuValAspGlyValArgGlnGly-----ThrAspThrGlyHis---	114
QY	352	GCATGAATACCGGTTTCATGCCCTCTGGCGGCCATGAGCGTATGAGTTATCAG	411
DB	115	---LeuAsnGlyThrPheLeuAspPro---AlaLeuIleLysArgValGluIleVal	132
QY	412	GGCCCATGTCACATGATGCTCTGATCGATGGCGGCTGTGGTGAATATCATTACC	471
DB	133	GlyProSerAlaLeuLeuTyrGlySerGlyAlaLeuGlyGlyValIleSerTyrAsp	152
QY	472	AGAAAGATGCAGACAAATGGCTCTCTCCGTCATAGCGGCTGAATCTGCAGAAAGC	531
DB	153	-----ValAspAlaLysAspLeuLeuGlnGlu---	161
QY	532	AACAAATGGGTAAACAGCAGCCAGTTTAATTTCTGAGCAGTGGTCCCTTGTGATG	591
DB	162	-----GlyGlnSerSerGlyPheArgValPheGlyThrGlyGlyThrGlyAsp	178
QY	592	TCGTGAGCGCTG-----CAGTACCGCGTACACACACAGCGTCAGCGTTTCAT	645
DB	179	SerLeuGlyLeuGlyAlaSerAlaPheGlyArgThrGluAsnLeuAspGlyIleVal	198
QY	646	ACATCACTGAGCGATACACAGCAGCGCT-----ATTCCATTATCCACGAG	693
DB	199	TrpSerSerArgAspArgGlyAspLeuArgGlnSerAsnGlyGluThrAlaProAsn	218

QY	694	TCACAGATTATTAATCTTGGTGGCAGCTCTTCACTGGAAGGCGTCGAGCAGGAT	747
DB	219	GluSerIleAsnAsnMetLeuAlaLysGlyThrTrpGlnIleAspSerAlaGlnSerLeu	238
QY	748	-----GTGCTCTGGTTTGTATATGATGATGATGATGATGATGATGATGATG	777
DB	239	SerGlyLeuValArgTyrTyrAsnAsnAspAlaArgGluProLysAsnProGlnThrVal	258
QY	778	CAGCGTTATGATAACCGGAGTGGCAACTGGGGAGTCTGACGGGGGATATGACCGG	837
DB	259	GluAlaSerAspSerSerAsnProMetValAspArgSerThrIleGlnArgAspAlaGln	278
QY	838	CTGCGCTATGAGCGAAACAAATTTTCACTGGCTGATCATCATCTTCCACCTTCGGA	897
DB	279	LeuSerTyrLys-----LeuAlaProGlnGlyAsnAsp	289
QY	898	TGG-----AAATCGTATCTGAACCTGGAACGAGACAGAAAAATAAGTCTGAGCTT	948
DB	290	TrpLeuAsnAlaAspAlaLysIleTyrTrpSerGluValArg-----	303
QY	949	GTACGCGAGTGTACTGAACGCGCAAAATGGGGCTTCCCGTCAGCCGCGGAG	1002
DB	304	-----IleAsnAlaGlnAsnThrGlySerSerGlyGluTyrArgGluGlnIle	319
QY	1003	CTTAAGGAATCGAACCTTATCTCTGAATTCATTTACTGCTTACCCTCTGGGAGAA	1062
DB	320	ThrLysGlyAlaArgLeuGluAsnArgSerThrLeuPheAlaAspSerPheAlaSerHis	339
QY	1063	CTGTTACGTTAGGGGGGAGTTCAGAGCTGCTCCATGAAGAACGAGTTCCTTGGC	1122
DB	340	LeuLeuThrTyrGlyGlyGluTyrTyrArgGlnGluGlnHisProGlyGly-----	357
QY	1123	AGCACAGTCAAACTTCCGCGCAGAAAGC-----TGGTCGGTATTGTCGAGGAT	1173
DB	358	ThrThrGly-----PheProGlnAlaLysIleAspPheSerSerGlyTyrLeuGlnAsp	375
QY	1174	GAGTGGCATCTACGGAT---GCATTTGCTGACTGGCGGCGACCGCTATGAACATCAT	1230
DB	376	GluIleThrLeuArgAspLeuProIleThrLeuLeuGlyGlyThrArgTyrAspSerTyr	395
QY	1231	GAGCAATTCGGGGACACTTCAGTCCGCGCATATCTGCTGCGGATG---GCAGAT	1287
DB	396	ArgGlySerSerAspGlyTyrLys-----AspValAspAlaAsp	408
QY	1288	GCCTGGACGCTGAAAGCGGTGTGACC-----	1314
DB	409	LysTrpSerSerArgAlaGlyMetThrIleAsnProThrAsnTrpLeuMetLeuPheGly	428
QY	1315	-----ACGGGATATAAGCAGCCAGCAGTGGGCGAGCTACATAAAGGATTAGTGT	1365
DB	429	SerTyrAlaGlnAlaPheArgAlaProThrMetGlyGluMetTyrAsnAspSerLysHis	448
QY	1366	GTGTGCC---GGGCGAGGAGAAACAAATCTA---CTTGGTAAACCCCGACCTGAAGCGG	1419
DB	449	PheSerIleGlyArgPheTyrThrAsnTyrTrpValProAsnProAsnLeuArgProGlu	468
QY	1420	-----GAGAGCTCAGTATGAGGCTGGGTGTATTACGAT-----AAC	1458
DB	469	ThrAsnGluThrGlnGluTyrGlyPheGlyLeuArgPheAspLeuMetLeuSerAsn	488
QY	1459	CCGCGCGTCTGAATGCCAATGTCACAGGTTTATGACT-----GACTTCTCCAAC	1509
DB	489	AspAla---LeuGluPheLysAlaSerTyrPheAspThrLysAlaLysAspTyrIleSer	507
QY	1510	AAGATTGCTCTTATTCATAAATGATAACCAATAGCTATGTAAACCGGAAAGGCC	1569
DB	508	ThrThrValAspPheAlaAlaThrThrMetSerTyrAsnValProAsnAlaLys---	526
QY	1570	CGGTTGCGCGGTGGAAATTTGCGCGCACATTGCGCGCTGTGTCAGAGGATGTC	1623
DB	527	-----IleTrpGlyTrpAspValMetThr	534


```

QY 1624 -----ACGCTGTCACTGAATTACACCTGGACCCGAAGTGAACAACGTGAT 1668
Db 535 LysTyrThrThrAspLeuPheSerLeuAspValAlaTyrAsnArgThrArgGlyAsp 554
QY 1669 GGTGATAACAAGGTGCGCCGCTGATTTATACCCCTGAACACATGGTGAATGCGAAACTG 1728
Db 555 ThrAspThr--GlyGluTyrIleSerSerIleAsnProAspThrValThrSerThrLeu 573
QY 1729 AACTGGCAGATCACCGAAGGTGGCATCA-----TGGCTGGGTGCCCGTTATCGCGGG 1782
Db 574 AsnIleProIleAlaHisSerGlyPheSerValGlyTrpValGlyThr---PheAlaAsp 592
QY 1783 AAACACACACCGTTTCAACCCAGAAATTTTCGTCACTGACGCGTGTACAGAAAGAAAGTGTAT 1842
Db 593 ArgSerThrHisIleSerSerSerTyrSerLys----- 603
QY 1843 GATGAGAAGAGAGATACCTGAAGCCCTGGACGCTGGTGGATGCGAGTCTCTGCTGGAG 1902
Db 604 -----GlnProGlyTyrGlyValAsnAspPheTyrValSerTyrGln 617
QY 1903 ATGACGGATGCCCTG-----ACGCTGAATGCTGCGGTGAATAACCTGCTCAACAAG 1953
Db 618 GlyGlnGlnAlaLeuLysGlyMetThrThrThrLeuValLeuGlyAsnAlaPheAspLys 637
QY 1954 GATTAC 1959
Db 638 GluTyr 639

```

RESULT 5

```

; Sequence 77625, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chisen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77625

```

```

; LENGTH: 698
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-10-282-122A-77625
Alignment Scores:
Pred. No.: 4,34e-24 Length: 698
Score: 371.00 Matches: 176
Percent Similarity: 41.59% Conservative: 143
Best Local Similarity: 22.95% Mismatches: 292
Query Match: 9,94% Indels: 156
DB: 12 Gaps: 35
US-10-625-972-4 (1-2091) x US-10-282-122A-77625 (1-698)
QY 16 CTGCTTCGCTAGTATCCCTGCTCGCATTTTCAGCC---AGCAGCATAGTCTGCTGCA 72
Db 12 LeuSerSerAlaIleMetLeuAlaValProSerAlaTyrAlaAspAspTyrAlaSer 31
QY 73 GAGGATGTGATGATTCCTCGCATCCGCTATGAGAAAAGCTGACTAACGCGCGCCG 132
Db 32 PheAspGluValValSerThrArgLeuAsnThrGlnIleThrAspThrAlaAla 51
QY 133 AGTGTTCCTGATAGCCAGGAGGATTCAGTCCAGCCAGTACCCAGCATCTGGCGGAG 192
Db 52 SerValAlaValIleAsnAlaSerAspIleGluGlnGlnMetAlaGluAspIleGluGly 71
QY 193 GCTCTGAGATCAGTAGAGGCTGTGGATGTTGAAAGTGGTACGGGTAAACCCGAGGCGTG 252
Db 72 LeuPheLysTyrThrProGlyValThrLeuThrThrAsnSer---ArgGlnGlyValGln 90
QY 253 GAAATCAGCATCCGAGGATGCCAGCCAGTACACGCTACACGCTGATCTACTGATTCGTTTCGT 312
Db 91 GlyIleAsnIleArgGlyIleGluGlyAsnArgIleLysValIleValAspGlyValAla 110
QY 313 CAGGCGGGAAGCAGTACGCTGACTCCCAACGGTTTTTCTGCCATGAATACCCGGTTCATG 372
Db 111 GlnProAsnGlnPheAsp-----SerGlyAsnSerPheLeuAsnSerSerArgVal 127
QY 373 CCC---CCTCTGGCCGCCATAGAGCTATTAGGTTATCAGGGGCGCGATGTCACACTG 429
Db 128 AspIleAspThrAspMetValLysSerValGluIleValLysGlyAlaAlaSerSerLeu 147
QY 430 TATGGCTCTGATCGATGGCGGTGGTGAATATCATTCACAGAAAGCAAAATGCGACAAA 489
Db 148 GlnGlySerAspAlaIleGlyIleValAlaPheGluThrLysAspProAlaAsp--- 166
QY 490 TGGCTCTCTTCCTCAATCGAGGCTGAATCTGCAGAAAGCAAAATGCGGTAAACAGC 549
Db 167 -----IleLeuLysGlyArgAsnMet-----GlyGlyTyr 176
QY 550 AGCAGTTTAAATTTCTGGACAGTGTCCCTTGTGGATGATTCCTCAGCTCGAGGTA 609
Db 177 AlaLysLeuAsnTyrSerSerAspLysThrPheSerGluSerIleAlaLeuAlaAsn 196
QY 610 CGCGGT-----AGCACAAACAGCGTCAGGTTTCATCGGTACCA 648
Db 197 LysSerGlyAspLeuGluSerLeuValAlaTyrThrArgAspGlyGlnGluIleGln 216
QY 649 TCACTGAGCATACAGCAGCAGCGGTATTCCTTATCCACGAGGTCACAGAATTATAT 708
Db 217 AsnPheGlySer-----ProAspGlnGlnAspAsnAlaAsnAsn 230
QY 709 CTTGGTGCAGTCTTGACTGGAAGCGTGGAGCAGGATGTGCTCTGGTTTGTATATGAT 768
Db 231 LeuLeuValLysLeuGlnTyrGlnLeuAsnProLysHisArgLeuGluPheSerGlyAsn 250
QY 769 ACCACCCGCGAGCTTATGATAACCGGATGGCAACTGGGGAGTCTG-----ACGGGG 822
Db 251 TyrIleArgAsnLysAsnAspLeuGluAsnLeuGluPheSerGlyTyrLysAsnAlaSer 270
QY 823 GGATATGACCGGACCTGCGCTATGAGCGAAACAAATTTACAGTGGCTGATCATCTACT 882

```


Db 118 ValArgArgGlyPheGlyAlaAsnArg-----AspGlySerIleMetThr 132
 Qy 316 GCGGACACAGTACGCTGACCTCCAAACGGTCTTCTGCGCATGATACCGGTTTCATGCC 375
 Db 133 AsnGlyLeuArgThrValLeuProArgSerPheAsn----- 144
 Qy 376 CCTCGCGCCATTGACGCTTATGAGTTATCAGGGGGCCGATGTCCACACTGTATGGC 435
 Db 145 -----AlaAlaThrGluArgValGluValLeuLysGlyProAlaSerThrLeuTy-Gly 162
 Qy 436 TCTGATGCGATGGCGGTGGTGAATATACATACAGAAAGATGCAGACAAATGGCTC 495
 Db 163 IleLeuAspProGlyGlyLeuIleAsnValThrLysArgProGluLysThrPheHis 182
 Qy 496 TCTTCGTCATGACGGCTGAATCTGCAGGAAAGCAAAATGGGGTAAACAGACGCCAG 555
 Db 183 GlySerValSerAla-----ThrSerSerPheGlyGlyThrGlyGln 198
 Qy 556 TTTAATTTCTGGAGCAGTGGTCCCTTGTGATGATCTCTGCAGCTGCAGGTACGGGT 615
 Db 199 LeuAspIle-----ThrGlyProIleGluGlyThrGlnLeuAlaTyArgLeuThrGly 216
 Qy 616 AGCACACAACAGCGTCAG-----GGT 636
 Db 217 GluValGlnAspGluAspTyTrpArgAsnPheGlyLysGluArgSerThrPheIleAla 236
 Qy 637 TCATCGGTACATCATCTAGCGATACACGACGCGCTATCTCTTATCC----- 687
 Db 237 ProSerLeuThrTrpPheGlyAspAsnAlaThrValThrMetLeuTySerHisArgAsp 256
 Qy 688 -----ACGGATCACACAATATATAAT 708
 Db 257 TyrLysThrProPheAspArgGlyThrIlePheAspLeuThrLysGlnProValAsn 276
 Qy 709 CTTGTGTGCACGCTCT----- 723
 Db 277 ValAspArgLysIleArgPheAspGluProPheAsnIleThrAspGlyGlnSerAspLeu 296
 Qy 724 -----GATCGAAGCGTCGAGCAGGATGCTCTGTTTCATATGGAT 768
 Db 297 AlaGlnLeuAsnAlaGluTyHisLeuAsnSerGlnTrpThrAlaArgPheAspTySer 316
 Qy 769 ACCACCCGCGACGCTAT---GATAACCGG-----GATGGCAACATG 807
 Db 317 TyrSerGlnAspLysTySerAspAsnGlnAlaArgValThrAlaTyAspAlaThrThr 336
 Qy 808 GGGAGTCTGACGGGGGATATGACCGACC----- 837
 Db 337 GlyThrLeuThrArgArgValAspAlaThrGlnGlySerThrGlnArgMetHisAlaThr 356
 Qy 838 ---CTGCGCTATGACGGAACAAATTTTCAGCTGGCTATGATCATCTTTCACCTTCGA 894
 Db 357 ArgAlaAspLeuGlnGlyAsnValAspIleAlaGlyPheTyAsnGluIleLeuGlyGly 376
 Qy 895 ACATGGAAATCGTATCTGAACCTGGAACGACAGACAGAAATAAAGTCGTGAGCTGTACGC 954
 Db 377 ValSerTyGluTy-Tyr-----AspLeuLeuArg 386
 Qy 955 AGTGACTGAACGCGACAAATGGGGCTTGCGGCTCAGCGCGGAGCTTAAGCAATCG 1014
 Db 387 ThrAspMetIleArgCysLys-----LysAlaLysAspPhe 398
 Qy 1015 AACCTTATCTGAATTCATCTACTGCTTACCCTCTGGGAGAACTCTCATCTGTTACGGTG 1074
 Db 399 Asn-----IleTyAsnProVal----- 404
 Qy 1075 GGGGGGAGTTTCAGAGCTGCTCCATGAAGACGAGGTTCCTTCGCAGCAGCAGGTGAA 1134
 Db 405 -----TyrGlyAsnThrSerLysCysThrThrValSerAlaSerAspSerAspGln 421
 Qy 1135 ACTTTCCGCGCAAAAGCTGGTCGCTATTGCTGAGGATGAGTGCATCTCACGATCCA 1194

Db 422 ThrIleLysGlnGluAsnTyrSerAlaTyAlaGlnAspAlaLeuTyLysThrAspAsn 441
 Qy 1195 CTTGCGCTGACTGCGGCGACCGCTATGAACATCATGACCAATTCGGGCGACACTTCAGT 1254
 Db 442 TrpIleAlaValAlaGlyIleArgTyGlnTyTrpThrGlnTyAlaGlyLysGlyArg 461
 Qy 1255 CCGCGTCGATATCTGCTCGGATGCGCAGATGCGTGGACGCTGAAGCGCGTGTGACC 1314
 Db 462 ProPheAsnValAsnThrAspSerArgAspGluGlnTrpIle----- 475
 Qy 1315 ACGGATATAAGGACCCAGAAATGGGCGAGCTACATAAAGGATTAGTGTGTGTCGGG 1374
 Db 476 -----ProLysLeuGlyLeuValTyLysLeuThrProSerValSer--- 489
 Qy 1375 CAGGAAAAACAATCTACTTGTGAAC----- 1401
 Db 490 -----LeuPheAlaAsnTySerGlnThrPheMetProGlnSerSerIle 504
 Qy 1402 -----CCCGACCTGAAGCGGAGAGAGCGTCACTTATGAGCTCGGGTGTAT 1449
 Db 505 AlaSerTyIleGlyAspLeuProProGluSerSerAsnAlaTyGluValGlyAlaLys 524
 Qy 1450 TAGCATAAACCCCGCGCTGCAATGCCAATGCACAGGTTTATGACTGACTTCTCCAAC 1509
 Db 525 PheGluLeuPheAspGlyIleThrAlaAspIleAlaLeuPheAspIleHisLysArgAsn 544
 Qy 1510 AAGATTGCTCTTATTCATTAATGATAACCAATAGCTATGTAACACGCGAAAGGCC 1569
 Db 545 ValLeuTyThrGluSerIleGlyAspGluThrIleAla---LysThrAlaGlyArgVal 563
 Qy 1570 CGTTTCACAGCTGTGGAA-----TTTGGCGGCACATTCGCGTGTGGTCAGAGGATGC 1623
 Db 564 ArgSerArgGlyValGluValAspLeuAlaGlyAlaLeu-----ThrGluAsnIle 580
 Qy 1624 ACCTGTCACTGAATTACACCTGCGACCGCAAGTGA---CAACGTGATGGTATACAAA 1680
 Db 581 AsnIleAlaSerTyGlyTyThrAspAlaLysValLeuGluAspProAspTyAla 600
 Qy 1681 GGTGCGCGCTGATGTTATACCTCCTGAACACATGCTGAATCGCAAACTGAACCTGCAGATC 1740
 Db 601 GlyLysProLeuProAsnValProArgHisThrGlySerLeuPheLeuThrTyAspIle 620
 Qy 1741 ACCGAAGAGGTGCATCA-----TGCTGGGTGCGCGTTATCGCGGGAACACCA 1791
 Db 621 HisAsnMetProGlyAsnAsnThrLeuThrPheGlyGlyGlyHisGlyValSerArg 640
 Qy 1792 CGTTTCACCCAGAAATTATTCCTCACTGACGCGCTGACAGAAAGTGTATGATGAGAAA 1851
 Db 641 ArgSerAlaThrAsnGlyAlaAspTy----- 649
 Qy 1852 GGAAGAATACCTGAAAGCCTGACGCGTGTGGATGCGAGGTCTGTCTGGGAAGATG----- 1905
 Db 650 -----TyrLeuProGlyTyPheValAlaAspAlaPheAlaAlaTyLysMetLysLeu 667
 Qy 1906 ACGGATGCGCTGACGTGAATGCTGCGGTCAATAACCTCTCAACAGGATTACAGTGAC 1965
 Db 668 GlnTyProValThrLeuGlnLeuAsnValLysAsnLeuPheAspLysThrTyThr 687
 Qy 1966 GTGAGCGCTG 1974
 Db 588 SerSerIle 690

RESULT 7

US-10-238-075-527
 ; Sequence 527, Application US/10238075
 ; Publication No. US20030148324A1
 ; GENERAL INFORMATION:
 ; APPLICANT: I.N.S.E.R.M.
 ; TITLE OF INVENTION: Polynucleotides which are of nature B2/D- A- and which are isolat
 ; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of thei
 ; FILE REFERENCE: BLANDINE
 ; CURRENT APPLICATION NUMBER: US/10/238,075
 ; CURRENT FILING DATE: 2002-09-10

Db 543 -----AsnArgThrGluSerMetAspAlaSerSerArgGluTrpLeuGlySerGlyAsn 560
 Qy 1702 CTTGAACACATGTTGAATGCGAAACTGAATGCGAGATCACCGAAGAGTGGCATATGG 1761
 Db 561 ProAspThrLeuIleSer-----AspIleSerIlePro 571
 Qy 1762 CTGGTGGCCCTG-----TATCGCGGGAACACACACCGTTTACCCAGATTTATCTCA 1815
 Db 572 ValGlyHisArgGlyValTyrAlaGlyTrpArg-----AlaGlu 584
 Qy 1816 CTGAGCGCTGACAGAGAAAGTGTATGATGAGAAAGAGAA-----TACCTGAAAGCCTGG 1872
 Db 585 LeuSerAlaSerAlaThrHisVal-----LysLysGlyAspProHisGlnAlaGlyTyr 602
 Qy 1873 ACGTGGTGGATGAGGTTCTGTCTGAGAGATGACGATGCGCTGAAATGCTGG 1932
 Db 603 ThrIleHisSerPheSerLeuSerTyrLysProValSerValLysGlyPheGluAlaSer 622
 Qy 1933 GTG-----AATAACCTCTCAACAGAGATTACAGTGACGTGACGTGACGTGCGCGT 1986
 Db 623 ValThrLeuAspSerAlaPheAsnLys-----LeuAlaMetAsnGly 636
 Qy 1987 AAGAGTACGCTGTATGCGCGTGTATCTTCAGAGCGGATCATCAACAGGATAT 2043
 Db 637 LysGlyValProLeuSerGly-----ArgThrValSerLeuTyrThrArgTyr 652

RESULT 8

US-10-282-122A-77578
 ; Sequence 77578, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Cart, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 77578
 ; LENGTH: 713
 ; TYPE: PR1

ORGANISM: Vibrio cholerae
 US-10-282-122A-77578
 Alignment Scores:
 Pred. No.: 5,57e-20 Length: 713
 Score: 325.50 Matches: 193
 Percent Similarity: 38.43% Conservative: 116
 Best Local Similarity: 24.00% Mismatches: 264
 Query Match: 8.72% Indels: 231
 DB: 12 Gaps: 48
 US-10-625-972-4 (1-2091) x US-10-282-122A-77578 (1-713)
 Qy 1 ATGCGAATAACCACTCTGGCTTCGCTAGTCAATTCCTGCTCGGATTTTCAGCCAGCAGC 60
 Db 1 MetLysLeuSerProValSerAlaAlaValLeuSerValLeu-----AlaAlaGly 17
 Qy 61 ATAGCTGCTGCAGAG-----GATGTGATGATGCTTCGCGCATCCGCGC 102
 Db 18 PheAlaHisAlaGluThrGluProSerHisTyrGluGluValValValAlaAsnArg 37
 Qy 103 TATGAGAAAAGCTGACTAAGCAGCGCCAGCTGTTCTGTGATTAGCCAGAGGAATG 162
 Db 38 IleGluGlnProLeuSerGluValAlaGlySerValAlaValLeuGluGluThrLeu 57
 Qy 163 CAGTCCAGCCAGTACCAGCATCTGCGGAGGCTCTGAGATCAGTAGAGGGTGTGATGTT 222
 Db 58 GluLysGlnGlyLysThrGluLeuTyrAspAlaLeuAsnGlnGluProGlyValSerVal 77
 Qy 223 GAAAGTGTACGGGTAAACCCGAGGGCTGGAATCAGCATCCGAGGAATCCAGCCAGT 282
 Db 78 ThrGlyGlyAlaGlyArgPro-----GlnAsnIleThrIleArgGlyMetThrGlyAsn 95
 Qy 283 TACACGCTGATCTACTGATTGATGTTCTGCTAGGGC-----GGAAGCAGTGCAGT 333
 Db 96 ArgIleAlaIleValArgAspGlyIleGlnSerAlaAspGlyTyrGlyAlaAlaAspIle 115
 Qy 334 ACTCCC-----AACGGTTTTCTGCCATGAATACCGGGTTCATGCCCTCTG 381
 Db 116 AsnAspLysTyrGlyArgAsnThrPheSerLeuSerAsn----- 128
 Qy 382 GCGCGCATGAGCGTATTGAGGTTATCAGGGGCGCATGCCACACTGTATGCTCTGAT 441
 Db 129 -----ValLysGlnIleGlnValValLysGlyAlaSerSerThrLeuTyrGlySerGly 146
 Qy 442 GCGATGGCGGTGTGTGATATATCATACAGAAAGAAATGCAGACAAATGGCTCTCTCC 501
 Db 147 AlaIleGlyGlyValVal-----IleIleGluSerLysAlaProGluAspTyrLeuTyrHis 165
 Qy 502 -----GTCAATGCGGGCTGAATCTCAGGAA-----AGCAACAATCG--GGT 543
 Db 166 ArgAspTyrTyrValAspAlaAlaLeuThrTyrSerGlyIleSerAsnArgIleGlnGly 185
 Qy 544 AACAGCAGC-----CAGTTTAAATTTCTGG 567
 Db 186 AsnHisAlaLeuAlaMetArgHisGlyAspGlyGluAlaLeuLeuThrIleAspTyrTrp 205
 Qy 568 AGCAGTGGTCCCTTGTGGATGATCTGTGAGCTCAGGTACGCGGTAGCACACACAG 627
 Db 205 ----- 205
 Qy 628 CGTCAGGTTTCATCGGTTCACATCAGTACGATACAGAGGACGCGGTATTCCTTATCCC 687
 Db 206 ---GlnGlyGluThrArgAsnPheAsnGlnAspLeu-----TyrAsn 219
 Qy 688 ACGAGTCAAGAAATATATCTGTGTGACGCTTGTGAGAGCGGTGCGGACGAGAT 747
 Db 220 ArgGluValAspGlyTyrAsnLeuGlyPheSerHisHisTyr----- 233
 Qy 748 GTGCTCTGGTTTGTAT-----ATGGATACACCGCGGAGCGTTATGATAAC--- 792
 Db 234 -----TrpLeuAsnAspAlaLeuArgLeuLysThrHisLeuGluTyrPheAspAspTyr 251

QY 793 -----CGGATGGCAACTGGG-----AGTCTGACGGG 822
 Db 252 AlaLysArgGluGlyThrSerSerIleGlnLysAspLysTrpAspLeuValSer 271
 QY 823 GGATATGACCGGACCTCGCTATGAGCAACAA-----ATTTCAGCTGGCTAT 873
 Db 272 PheTyrGlu-----TyrGlnArgSerGlnThrArgLeuAlaSerValGlyAla 287
 QY 874 GATCATATTTCACCTTCGGAACATGAAATCGTATCTGAAC-----TGGAAC 921
 Db 288 AspTyrThrAlaAsnLeu-----SerTrpMetAspThrLeuGluGlyLysPheTyrTrpArg 306
 QY 922 GAGACAGAAATAAAGCTCGTAGCTGTGACGAGGTGTACTGAAGCGGCACAAATGGGG 981
 Db 307 SerThrGluAsnIleThrGlnThrAsnArgLeuMetAlaAsnAspArgSerGlyAlaGly 326
 QY 982 CTTGCGGTCAGCGCGGAGCTTAAGGAATCGAACCTT-----1020
 Db 327 IleLeuSerTyrArgArgGluLeuArgAspGluGlyPheAsnAspGluAlaLeuGlyAla 346
 QY 1021 ATCCTCAATTCTACTGCTTACCTCCCTCGGAGAA-----TCTCATCTG 1065
 Db 347 ThrLeuAsnIleGlnLysGluTrpGlnGlnGlyGluTrpLeuHisGlnPheAlaTyrGly 366
 QY 1066 GTTACGCTGGGGGC---GAGTTTCAG-----AGTCGTCC 1098
 Db 367 MetSerValAspGlyHisAspTyrGlnArgProLysSerIleArgArgMetGluSerSer 386
 QY 1099 ATCAAGACGAGTTCCTCTCCACACAGGTGAACCTTC-----CGGAGAAA 1149
 Db 387 GlyAspAsp-----LeuGlnAlaAspGluProPheAlaProAlaArgGluTyr 402
 QY 1150 AGTCGTCTGCTATTGCTGAGGATGAGTGGCATCTCACGATGCATCTCGCTGACTGCG 1209
 Db 403 ArgPheGlyValTyrGlyGlnAspAsnLeuLeuGlyAsp---TrpThrLeuAlaAla 421
 QY 1210 GGACGCGCTATGAA-----CATCATGACCAATCGGGGACAC 1248
 Db 422 GlyLeuArgPheAspAlaGlnLysLeuSerProLysAsnThrAspArgIleHisGlyTyr 441
 QY 1249 -----TTTCAGTCGCGTGCATATCTGCTCGGATG 1281
 Db 442 LysValValThrMetGlySerSerGluTrpSerProSerAlaSerIleSerTyrGlnTrp 461
 QY 1282 GCAGATGCTGCAGCTGAAAGCGGTGTGACCGGATATACCGACCCAGAAATGGGG 1341
 Db 462 HisProGluTrpAsnThrTyrLeuSerTyrAsnHisGlyPheArgAlaProSerTyrAsp 481
 QY 1342 CAGCTACATAAAGGATTAGTGTGTCTCGGAGGAAACAAATCTACTGGTAAC 1401
 Db 482 LysAlaTyr---GlyAlaSerAspHisSerPheValProLeuThrProPheIleLeuLys 500
 QY 1402 CCGGAC-----CTGAAGCGGAAGAGCGTCAGTTATGAGCTGGGTGATTACGAT 1455
 Db 501 ProAsnAsnLysLeuAlaGluThrSerAspSerPheGluLeuGlySerLysTyrAsp 520
 QY 1456 AACCCCGCGTCTGATCCATGTCACAGGTTTATGACGACTTCTCCAAC-----1509
 Db 521 Asn---GlyGlnThrGlnPheTyrValAlaValPheTyrSerIlePheAspAsnPheIle 539
 QY 1510 -----AAGATTCTCTTCTTATCCATAATGATGAACACC-----AATAGC 1548
 Db 540 AspValLysGlnValGlyTyr-----AspAsnAlaThrGlySerValIleGlnGln 556
 QY 1549 TATGTAAACAGCGGAAGCCCGGTGACGCTGTGGAATTTGCCGCGACATTCGCGCTG 1608
 Db 557 TyrGlnAsnIleAlaGlyValLysThrTyrGlyAlaGluMetSerValMetHisArgLeu 576
 QY 1609 -----TGCTCAGAGGATGTCAGCTGTCTCACTGAATTTACACCTGACCGCAAGTAA 1659
 Db 577 AspAspArgTyrSer-----ValGluAsnLysLeuGlyTyrVal-----Asp 590
 QY 1660 CAACGTGATGGTGATAACAA-----GGTGGCGCG 1689

Db 591 GlyLysAspGlyGluAsnGlnTyrValArgThrLeuThrProLeuGluGlySerValGln 610
 QY 1690 CTGAGTTATACCCCTGAACACATG---GTGAATCGGAACATGAACTGGCAGATCACCGAA 1746
 Db 611 LeuAsnTyrGlnArgGluArgTrpAspAlaTyrSerArgLeuAsnTrp-----626
 QY 1747 GAGTGGCATCATGCTGGTGGCGCTTATCGCGGAAACACACGCTTTTACCAGAAAT 1806
 Db 627 -----AlaSerAlaMetSerArgValProThrCysThr-----637
 QY 1807 TATTGCTACTGAGCGGTGTACAGAAAGTGTATCATGAGAAAGAGATACCTGAAA 1866
 Db 638 -----ThrGluGlnGlyLysGluThrGlu 645
 QY 1867 -----GCCTGGACGGTGGTGATGCGAGTCTGTCTGGAGATGACGGAT 1911
 Db 646 CysAlaThrThrThrGlyTrpValSerTrpAspIleGlyLeuAsnTyrGlnTrpAsnAla 665
 QY 1912 GCCCTGACGCTGAATGCTGGGTGAATACTCTCAACAAAGGATTAACAGT-----1962
 Db 666 GlnLeuSerAlaSerPheAsnValValAsnLeuLeuAspArgGluTyrThrArgTyrGln 685
 QY 1963 GAGCTGAGCTGTACAGTGGCGGT-----AAGAGTACGCTGTATGCC-----2004
 Db 686 AspVal-----AlaGlyValThrProSerAspThrLeuTyrSerThrGluPro 701
 QY 2005 GGTGATTACTTC 2016
 Db 702 GlyArgTyrPhe 705
 RESULT 9
 US-10-282-122A-48409
 ; Sequence 48409, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITEA 034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.


```

; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 48409
; LENGTH: 775
; TYPE: prt
; ORGANISM: Bacteroides fragilis
US-10-282-132A-48409

```

Alignment Scores:

Pred. No.: 1	Length: 775
Argument Scores:	Matches: 155
Arg. No.: 1	Score: 321.00
Percent Similarity: 39.3%	Conservative: 121
Best Local Similarity: 22.08%	Mismatches: 278
Query Match: 8.60%	Indels: 148
DB: 12	Gaps: 30

US-10-625-972-4 (1-2091) x US-10-282-122A-48409 (1-775)

QY	82	ATGATTGTCTGGCATCCGGCTAAGAGAAAAGCTGACTAACCGACGCCCGCAGTGTTCTT	141
Db	101	ValValValThrGlySerArgThrAlaArgProIleLysMetSerProValThrGln	120
QY	142	GTGATTAGCCAGGAGAAATTCAGTCCAGCCAGTACCACGATCTGGCGGAGGCTGTG	198
Db	121	ValLeuGlyGlyLeuValAspAlaGlyTyrSerAsnLeuGlnAlaLeuGln	140
QY	199	AGATCAGTAGACGGTGTGGATGTTCAAGATGCTACGGGTAAACCCGGA	255
Db	141	GlnGluThrProGlyLeuAsnIleGln	156
QY	256	ATCAGATCCGAGGAATCCGACCGATTCACGCTGATCTGATGATGGTGTTCGTACG	315
Db	157	IleSerMetGlnGlyLeuAspAlaArgHisValLeuPheLeuMetAspGlyGluArgMet	176
QY	316	GCGGAGGACGAGTACGTGACTCCCAAC	372
Db	177	ThrGly	190
QY	373	CCCCCTCTGGCCCATTTGAGCGTATTGAGGCTATTCAGGGGCGGATGCCACACGTAT	432
Db	191	LeuHisAlaIleAspArgValGluLeuValLysGlyAlaSerSerThrLeuTyr	208
QY	433	GGCTCTGATCCGATGGCGGTGTGGTGAATATCATACCAGGAATGCAGACAATGG	492
Db	209	GlySerArgAlaAlaGlyAlaValIleAsnLeuIleThrLysLys	227
QY	493	CTCTCTCCGTCAATGCAGCGGTGAATCTGCAGGAAGAACAAATGGGTAAACAGCAG	552
Db	228	LeuSerIleAspAlaGlyIle	240
QY	553	CAGTTTAATTTCTGGAGCAGTGGTCCC	597
Db	241	GluArgAsnTyrLysHisProGlnProLysAspPheLeuTyrMetPheGluGlnAsnAla	260
QY	598	AGCCTCAGGTACGGGTAGCACACACAGCGTCAGGGTTCATCGGTCCATCACTCAGC	657
Db	261	Asp	277
QY	658	GATACAGCAGCAGCGGTATTCCTTATCCACGGAGTCACAGAATTATATCTTGGTGA	717
Db	278	PheThrSerGlnThrAspValTyrSerGluSerAspAlaPheTyrMetTyrGlnAla	297
QY	718	CGTCTTGTGAGCGCTCGAGCAGAGATGTGCTCTGGTTT	759
Db	298	GluAsnAspLysLysValTyrThrLysGluAlaAsnProPheLeuProHisAspIleIle	317
QY	760		771
Db	318	ValValSerAsnAlaValArgProProMetGlyIleGluGlyLysGluHisIleThrVal	337
QY	772	ACCCGGCAGCGTTATGATAACCGGATGGCAACTCGGGAGTCTGACGGGGGA	825
Db	338	SerGlnLysLeuTyrThrAsnProAsnProAsnLeuSerValLeuValTyrGlySerSer	357

Qy	826	-----TATGACCGGACCCCTCGCTATGAGCGAAACAAA	858
Db	358	PhePheMetAsnThrTyrAspLeuIleGlnAspMetThrPheSerGlnAlaArgAsp---	376
Qy	859	ATTTCAGCTGGCTATGATCATACTTTCCACTTCGGGAACATGGAATCG-----	906
Db	377	TripThrAlaGlyThrLysValThrTyrHisValLysAspTppPheSerValThrGlySer	396
Qy	907	-----TATCTGAACCTGGAACGACAGACAGAAAATAAAGCTCGTGAGCTTGTA	951
Db	397	LeuHisAlaAspPheTyrAspArgPheLysArgHisGluArgIleAspLysArg----	414
Qy	952	CGCAGTGTACTGAAGCGGACAAATGGGGCTTGC CGGTACAGCGCGGAGCTTAAGGAA	1011
Db	415	-----GlnLysAspTyrGluSerSerIleTyrGlnProArgLeuThrValThr	430
Qy	1012	TCGAACCTTATCCTGAATTCATTACTGCTTACCCCTCTGGAGAATCTCATCTGGTTACG	1071
Db	431	SerAsnTyrPhe-----AsnGlyHisSerLeuIle	440
Qy	1072	GTGGGGGGCGAGTTTCAGAGCTCGTCATGAAGAAGCGAGTGTCTCTCCGACACAGGT	1131
Db	441	LeuGlyMetGluHisThrSerAspGluLeuThrSerAspArgPheSerGlyAsnAlaAsn	460
Qy	1132	-----CAAACTTTCGGCAGAAAAGCTGGTGGTATTTCCTCAGATGAGTGGCAT	1182
Db	461	HisAspLeuLysThrArgAlaLeuLysGluThrGluTyrPheLeuGlnAspGluTppThr	480
Qy	1183	CTCAGGATGCATCTGGCTGACTGGCGGAGCGGCTATGAACATCATGAGCAATTCGGG	1242
Db	481	IleAsnProArgTppMetIleSerValGlyIleArgThrAsnPheSerLysAlaPheGly	500
Qy	1243	GGACACTTCAGTCCGGCTGCATCTGCTGGATGTGGCAGATGCTCGACGCTGAA	1302
Db	501	PheMetGlyMetProLysValAlaAlaLysTyrSerProAspLysHisTppSerLeuArg	520
Qy	1303	GGCGGTGTACACGGGATATAAGCACCCAGAAATGGGCGAGTA-----	1347
Db	521	AlaAsnTyrSerMetGlyTyrArgSerProSerIleLysGluLeuPheAsnTppAsp	540
Qy	1348	CATAAAGGATTAAGTGGTGTCCGGCAGGGAAAAACAAATCTACTTGTATAACCCGAC	1407
Db	541	HisLeuGlyMetPheMetIleArg-----GlyAsnGluAsn	552
Qy	1408	CTGAAGCCGGAAGAGACGTCAGTTATGAGCTGGGTGGTATTACGATAACCCGCGGT	1467
Db	553	MetArgProGluLysAsnAsnTyrPheSerLeuGlyAlaGluTyrSerAsnAspArg---	571
Qy	1468	CTGAATGCCAATGTCACAGGTTTTATGACTGACTTCTCCAAAGAATGTGCTCT---	1521
Db	572	LeuPheValSerGlyThrAlaTyrGlyAsnTyrPheArgAspLysIleGluGlyValTpp	591
Qy	1522	-----TATTCCATAAATGATAACCAACVAGCTATGTAACAGCGGAAAGCCCGTTG	1575
Db	592	ArgIleTyrAspMetGlnTyrAsnPhe---GluTyrThrAsnLeuSerGlnArgGln	610
Qy	1576	CACGCTGTGAATTTGCCGCGACATTCGCCGTGGTGGTCAGAGGATGTCAGCTGTCACTG	1635
Db	611	LeuGlyLeuGlu-----ValGlnThrArgTppSerValLeuAspCysLeuThrLeu	627
Qy	1636	AATTACACCTGGACCGGAAGTGAACACACGTCATGGTGATTAACAAGGTGGCCGCTAGT	1695
Db	628	AsnGlyThrTyrSerPheValAspVal-----SerLysAsnLysGlyIleGlnValAsn	645
Qy	1696	TATACCCCTGAACACATGGTGAATGCCAAACTGAACCTGGCAGATCAACCGAAGAC---	1749
Db	646	ThrThrSerProHisAlaAlaThrAlaSerMetAsnTyrLysTyrMetLysLysAsnTyr	665
Qy	1750	-----GTGGCATCATGGCTGGGTCCCGCTTATCCGGGAAACACCA	1791
Db	656	ArgLeuAsnAlaValPheSerAlaSerTyrMetGlyGlyLys-----	679

172 QY CAGTACCACGATCGCGGAGGCTCTGAGATCAGTAGAGGCTGGATGTTGAAGTGGT 231
173 Db |||||
95 Db GlnPheGlnSerValLysGluAlaLeu----- 103
232 QY ACGGTAACACCGAGGCTGGAATCAGCATCCGAGGAATGCCAGCAGTTACACGGTG 291
233 Db |||||
104 Db --GlyTyrThrProGlyValThrValSerSerArgGlyAlaSerAsnThrTyrAspPhe 122
292 QY ATACTGATTGATGGTGTCTCAGCGGAGGAGCAGTACGTACTCCCAACGGTTT--- 348
293 Db |||||
123 Db ValIleIleArgGlyPheSerSerValGlyLeuSerGln-----AsnAsnTyrLeu 139
349 QY -----TCTGCCATGAATACCGGGTTCATGCCCTCTGCGCGCC 387
140 Db |||||
140 Db AspGlyLeuLysLeuGlnGlyAsnPheTyrAsnAspAlaValIleAspProTyr---Met 158
388 QY ATTGACGCTATTGAGTTATCAGGGGCGCATGTCCACACTGATGCTCTGATCGATG 447
389 Db |||||
159 Db LeuGluArgValGluLeuMetArgGlyProThrSerValLeuTyrGlyLysSerAsnPro 178
448 QY GCGGCTGTGTGAATATCATTACAGAAAGAACGAGCAAAATGGCTCTCTCCGTCAT 507
449 Db |||||
179 Db GlyGlyIleIleSerMetValSerLysArgProThrThrGluProLeuLysGluIleGln 198
508 QY GCAGGCTGAATCTGCAGGAAGCAACAATGGGTAAACAGCAGCGAGTTTAAATTCCTGG 567
509 Db |||||
199 Db PheLysMetGlyThrAspAsnLeuPheGlnThrGly-----PheAspPheSer 214
568 QY AGCAGTGGTCCCTTCTGGATGATCTCTGCACGCTCAGCCTCAGGTACGC----- 612
215 Db |||||
613 QY GTPAGCACACACAGCGTCCAGGCTTCATCGGTTCACATCCTGAGCGATACAGCGGCACG 672
232 Db |||||
673 QY CGTATTCCTTATCCACGAGGACGACAGATTAATATCTTGGTGACGCTTGAACGGAAG 732
242 Db |||||
733 QY GCGTCGAGCAGGATGCTCTGCTGTTT-----GATATGGATACCAACC 774
256 Db |||||
775 QY CGGACGCTTATGATTAACCGGAGTGGCAACTGGGGAGTCTGACGGGGGATATGACCGG 834
276 Db |||||
835 QY ACCCTG-----CGCTATGACGGAACAAATTT 861
294 Db |||||
862 QY TCAGCTGGCTAT-----GATCATACTTTC-----ACCTTCGGAACATGGAATCGTAT 909
314 Db |||||
910 QY CTGAAGTGAACGACAGACAGAAATAAGGTCGTGAGCTTGTACGAGTGTACTGAGCGC 969
333 Db |||||
970 QY GACAAATGGGGCTTCCCGTTCAGCGCGGAGCTTAAGGAATCGAACCTTATCTCGAAT 1029
349 Db |||||
1030 QY TCATTACTGCTTACC-----CCTCTGGGAGAA 1056
366 Db |||||
1057 QY ---TCTCATCTGCTGCGTGGGGCGAGTTTTCAGACTCGTCCATGAAAGACGGAGTT 1113
386 Db |||||
1114 QY GTCCTTGCACGACAC----- 1128

406 Db ThrPheGlySerAlaProSerIleAspLeuTyrAsnAsnTyrHisProGluTyrPheAla 425
1129 QY -----GGTGAACCTTTCCGG-----CAGAAAAAGCTGGTCCGATTGCTGTGAG 1170
426 Db |||||
426 Db PheGlyGlyAlaGluProTyrGlnMetAsnLysSerGlnThrGlyLeuTyrValGln 445
1171 QY GAT-----GAGTGGCATCTCACGGATGCTCCGCTGACTCGCGGACACCGCTATGAA 1224
446 Db |||||
1225 QY CATCATGACCAATTCGGGGGACACTTCAGTCCGGTGCATATCTGGTCTGGGATGTGGCA 1284
463 Db |||||
1285 QY GATCCCTGGACGCTGAAGCGGTGTGACCAACGGGATATAAGGCACCCAGAAATGGGGCAG 1344
467 Db ---AlaThrThrValArgGlnAsnSerThrThr-----ProThrGluGlyTyr 481
1345 QY CTACATAAA-----GGGATTAGTGGTGTCTCGGGCAG 1377
482 Db |||||
1378 QY GGAAAAACA-----AATCTACTT 1395
502 Db |||||
1396 QY GGTAAACCC-----GACCTGAAGCCGGAAGAGAGCGTCAAGTTATGAGCTGGGTGTAT 1449
522 Db |||||
1450 QY TACGATAAACCCCGCGTCTGAATGCAATGTCCACAGGT-----TTTATGACTGACTTC 1503
542 Db |||||
1504 QY TCACACAGATTTCTCTTATTCATTAATATGATAACCAAT-----AGCTATGTAAC 1557
561 Db |||||
1558 QY AGCGAAAGCCCGCTTGCACGCTGTGAATTTGCCGCGACATTCGCGTGTGGTCAGAG 1617
577 Db |||||
1618 QY GATGTCACGCTGCTCACTGAATTAACCTGACCCGGAAGTGA-----CAACGCTGATGCTGAT 1674
596 Db |||||
1675 QY AACAAAGGTGCGCGCTGAGTTATACCCCTGACACATCGTGAATCGCAAACTGAACCTGG 1734
616 Db |||||
1735 QY CAGATCACCGAAGAGGTGGCATCATCGCTG-----GGTCCCGCTTATCGCGG 1782
636 Db |||||
1783 QY AANAACACCGTTTCCACCAGAAATTTCTGCTACTGAGCGCTGTACAGAAAGATGAT 1842
656 Db |||||
1843 QY GATGAGAAAGGAGATATCCTGAAAGCCTGGACG-----GTGGTGGATGACGCTGTGTCG 1896
659 Db |||||
1897 QY TGAAGATGACG-----GATGCCCTGACCTGAATGCTCGGTGAAT 1938
679 Db |||||
1939 QY AACCTGCTCAACAAAGGATTAC 1959
697 Db |||||
703 Db AsnLeuLeuAspArgGluTyr 703

RESULT 12

US-10-282-122A-48639


```

; Sequence 48639, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48639
; LENGTH: 763
; TYPE: PRT
; ORGANISM: Bacteroides fragilis
; US-10-282-122A-48639

Alignment Scores:
Pred. No.: 9,298-18
Score: 161
Matches: 161
Conservative: 112
Mismatch: 334
Indels: 132
Gaps: 28
DB: 12

US-10-625-972-4 (1-2091) x US-10-282-122A-48639 (1-763)
QY 13 ACTCTGGCTCCGCTAGTCTCCCTGCTCGGATTTTCAGCCAGCAGCAGTCTGCTGCA 72
Db 91 ThrValSerAlaGluAlaThrSerValValHisPheGlnMetGluGluValSerPheThr 110
QY 73 GAGGATGTGATGATTTCTCGGCTATCGGCTATGAGAAAAGCTGACTAACCGCCGCC 132
Db 111 ThrAspGluValValSerAlaAsnA-GAsnGluValSerArgLysAlaAlaProVal 130
QY 133 AGTGTCTCTGATAGCAGGAGGAATTCAGTCCAGCCAGCAGTACACGATCTGGCGGAG 192
Db 131 ValValAsnValMetSerAlaLysLeuPheGluThrValAsnSerThrAspLeuAlaLys 150
QY 193 GCTCTGAGATCAGTAGGGTGTGGATGTTGAAGTGTGACGGGTAAACCGGAGGGCTG 252
Db 151 SerLeuAsnPheGlnSerGlyLeuArgValGluAsnAsnCysGlnAsnCysGlyPhePro 170

```

```

QY 253 GAAATCAGCATCCGAGGAATGCCAGCATTTACAGCTGATATCTGATGATGGTGTCT 312
Db 171 GlnValArgIleAsnGlyLeuGluGlyProTyrSerGlnIleLeuIle----- 186
QY 313 CAGGGCGGAAGCAGTGCAGTGCCTCCACAGCGTTTCTGCCATGAATACCGGTTCAAG 372
Db 187 -----AsnSerArgProIleIleSerAlaLeuSerGlyValTyrGlyLeuGluGln 203
QY 373 CCGCTCTCGCGCCCATTCAGGCTATTGAGGTTATCAGGGGCGCGATGTCACACTGTAT 432
Db 204 IleProValAsnMetIleGluArgValGluValValArgGlyGlyGlySerAlaLeuPhe 223
QY 433 GGCTCTGATCGGATGGCGGTGTGGTGAATATCATTCACAGAAAGNATCCACACAATGG 492
Db 224 GlyAlaAsnAlaValGlyGlyThrIleAsnIleIleThrLys----- 237
QY 493 CTCTCTTCCTCAATGCAGGCTGGAATCTGCAGGAAGCAACAATGGGTAAACAGCAG 552
Db 238 ---AspProIleAsnAsnSerPheGlnValAlaSerThrMetSerAsnMetAsnGlyLys 256
QY 553 CAGTTTAAATTTCTGGAGCAGTGTCCCTGTGGATGATCTGTGAGCTGTGAGGTACGC 612
Db 257 SerTyrGluGlnTyrMetGlyGly-----AsnValSerLeuValAla--- 270
QY 613 GTAGCACACACAGCGTTCAGGCTTCATCGGTACATCTGAGCATGATACAGCAGCAGC 672
Db 271 -----LysAspAsnSerTyrGlyIleAlaLeuTyrGluThrTyrArgAsn 285
QY 673 CGTATTCTTATCCACGAGTACACAGAAATATATCTTTGGTGCACCTCTGTACTGGAAG 732
Db 286 ArgAsnProTyrAspAlaAspGlyAspGlyPheSerGluLeuGlyLysLeu----- 302
QY 733 GCCTCGGAGCAGGATGCTCTGTTGATATGATACACCGCGGCGGCTTATGATAAC 792
Db 303 -----AsnMetAsnThrPheGlyMetArgAlaTyrTyr 313
QY 793 CGGATGGCACTGGGAGTCTGACGGGGGATATGACCGGAGCTTCCCTGCTATGACGGA 852
Db 314 ArgProAsnTyrPheSerArgIleAsnValGluTyrHisThrThrAsnGluPheArgArg 333
QY 853 -----AACAAA-----ATTCAGCTGGCTATGATCAT 879
Db 334 GlyGlyAsnLysPheAsnLeuGlnProHisGluAlaAspIleThrGluGlnThrLysHis 353
QY 880 ACTTTCACCTTCGGAACATCGAATCGTATCTGAACTGGAACAGACAGACAGAAATAAGGT 939
Db 354 IleIleAsnSerGlyGlyValSerTyrAspArgTyrTyrGlyGlyLysHisLys----- 371
QY 940 CGTGAGCTTGTACGAGTGTACTGAAGCGGCACAAATGGGGGCTTGCCTGACCGCGCG 999
Db 372 MetSerValTyrGlySerValGlnHisThrAspArgAsnSerTyrTyrGlyAlaGlnLys 391
QY 1000 GAGCTT-----AGGAATCGAACCTTATC----- 1023
Db 392 AspMetAsnAlaTyrGlyLysThrAsnAspLeuThrTyrValValGlyGlyMetTyrVal 411
QY 1024 -----CTGAATTCATTACTTACCCCTCTCGGAGAAATCTCATCTGTTACGGTGGG 1077
Db 412 GlyAsnMetAspArgCysLeuPheAlaProAla-----ThrPheThrGlyGly 427
QY 1078 GCGAGTTTACAGCTCGTCCATGAAAGACGCGGATTTCTCTTCCAGCAGCAGGTGAACT 1137
Db 428 ValGluTyrGlnSerAsnSerLeuHisAsp-----ValMetThrGlyTyrHisArgAsp 445
QY 1138 TTCGGCGAG-----AAAAGCTGCTCGGTATTTCTGAGGATGATGGCATCTCACCGAT 1191
Db 446 MetGlnGlnAspValArgIleAlaGlyGlyPheValGlnAsnGluTyrArgLeu--- 464
QY 1192 GCATCTGCGCTACCTCGGCGGCGCGCTATGAACATCATGAGCAATTCGGGGGACAC--- 1248
Db 465 ArgTyrThrMetLeuValGlyAlaArgLeuAspLysHisAsnLeuIle---AspHisPro 483

```


D	b	222	MetAspAlaIleAspValValArgGlyGlyAlaValArgTyrGlyProGlnSerVal	241
Q	y	448	GCGCGTGCTGGTAATCATATTACCAAGAAAGTGCAGAAAATCGCTCTTCOGTCAAT	507
D	b	242	GlyGlyValValAsnPheValThrArgAlaIleProGlnAspPhe-----GlyIleGlu	259
Q	y	508	CGAGGCTGAATCTGCAGAAAGCAACAATGGGGTAACACGCCAGCTTTAATTTCTGG	567
D	b	260	AlaGlyValGluGlyGlnLeuSer-----ProThrSerSerGlnAsnAsn-----	274
Q	y	568	AGCAGTGGTCCCCTGTGGATGATTTCTGTACGCTGCAGGTACGCGGTAGCACACAACAG	627
D	b	275	-----ProLysGlnThrHisAsnLeuMetValGlyGlyThrAlaAspAsn	289
Q	y	628	CGTCAGGGTTCATCGGTACATCATCAGCGATAGACGAGCAGCCGCTATTCCTTATCCC	687
D	b	290	GlyPheGlyThrAlaAsnLeu-----TyrSerGlyThrArgGlySer-----	303
Q	y	688	ACGGAGTTCAGAATAATAATCTGTGTGCACGCTCTTGACTGTGAAG-----CGCTCG	738
D	b	304	-----AspTrpArgGluHisSerAlaThr	311
Q	y	739	GAGCAGGATGTGCTCTGGTT-----GATATGGAT	769
D	b	312	ArgIleAspAspLeuMetLeuLysSerLysTyraLaProAspGluValHisThrPheAsn	331
Q	y	769	ACCACCGGCGAGTATTGATAACCGGATGGCAACTGGGAGTCTG-----	816
D	b	332	SerLeuLeuGlnTyrTyrAspGlyGluAlaAspMetProGlyGlyLeuSerArgAlaasp	351
Q	y	817	-----ACGGGGGATATGACCGGACCCCTCGGTATGAGCGA	852
D	b	352	TyrAspAlaAspArgTrpGlnSerThrArgProTyrAspArgPhe-----TrpGlyArg	369
Q	y	853	AACAAATT---TCAGCTGGCTAT-----GATCATACTTTCACCTTC	891
D	b	370	ArgLysLeuAlaSerLeuGlyTyrGlnPheGlnProAspSerGlnHisLysPheAsnIle	389
Q	y	892	---CGAACATCGTAATCTGTAATCTGAACTGAAACGACAGACAGAAAAATAAAGTCGTGAGCTT	948
D	b	390	GlnGlyPheTyrThrGlnThrLeuAsrGserGlyTyrLeuGlu-----GlnGlyLysArgile	408
Q	y	949	GTACGAGTGTACTGAAGCGCGCAAAATGGGGCTTCGCGTCACGCCGGGAGCTTAAG	1008
D	b	409	ThrLeuSer-----ProArgAsnTyrTipValArgLysGlyLeuProArg-----	423
Q	y	1009	GAATCGAACCTTATCTGTAATTCATTACTGCTTACCCCTCTGGGAGAATCT---CATCTG	1065
D	b	424	TyrSerGlnIlePheMet-----IleGlyProSerAlaHisGlu	436
Q	y	1066	GTTCACGTTGGGGCGAGTTTCAGAGCTGTCATGAAGACGGAGTGTCTTCGCCAGC	1125
D	b	437	ValGlyValGlyTyrArgTyrLeuAsnGluSerThrHisGluMetArgTyrTyrThrAla	456
Q	y	1126	ACAGTGAACCTTCCGGCAGAAAAGCTGGTGGTATTGTCAGGAT-----	1173
D	b	457	ThrSerSerGlyGlnLeuProSerGlySerSerProTyrAspArgAspThrArgSerGly	476
Q	y	1174	-----GAGTGGCATCTCACGGATGCACCT-----CCGCTGACT	1206
D	b	477	ThrGluAlaHisAlaTrpTyrLeuAspAspLysIleAspIleGlyAsnTrpThrIleThr	496
Q	y	1207	CGGGCAGCCGCTATGAACATCATGACCAATTC-----GGGGACAC	1248
D	b	497	ProGlyMetArgPheGluHisIleGluSerTyrGlnAsnAsnAlaIleThrGlyThrHis	516
Q	y	1249	TTCAGT-----CCCGCTGCATATCTGGTCTGGGATGTCGGAGAT	1287
D	b	517	GluGluValSerTyrAsnAlaProLeuProAlaLeuAsnValLeuTyrHisLeuThrAsp	536
Q	y	1288	GCCTGGAGCGCTGAAGCGGTGTGACACCGGGATATAAGCACCCAGANTGGGGCAGCTA	1347
D	b	537	SerTrpAsnLeuTyrAlaAsnThrGluGlySerPheGlyThrValGlnTyrSerGlnIle	556


```
Db 500 ProTyrPheSer-----TyrSerGluSerPheGluProAlaSerGlnThrAspAla 516
QY 1375 CAGGGAACAAATCTACTTGTGTAAACCCCGCTGAAGCCGGAAGAGAGCGTCAGTTAT 1434
Db 517 GlnGlyLysLeu-----PheSerProSerLysGlyLysGlnTyr 529
QY 1435 GAGCTGGGTGTATTAC-----GATAACCCCGCGGTCTGAATGCCAATCTCACA 1485
Db 530 GluAlaGlyValLysTyrValProAsnAspArgPro-----IleValValThr 545
QY 1486 GGT-----TTTATGACTGACTTCTCAACAAGATTGTCTTATTCCATAATGATAAC 1539
Db 546 GlyAlaLeuTyrGlnLeuThrLysThrAsnAsnLeuMetAla-----AspPro 561
QY 1540 ACCAATAGCTAT-----GTAACACGCGGAAGCCGGTTCACCGGTGTGAATTTGCC 1593
Db 562 AlaGlySerPhePheSerValGlnGlyGlyGluIleArgAlaArgGlyValGluLeuGlu 581
QY 1594 GGCACATTGCGCTGGTGCAGAGGATGTCACGCTGCTACTGAATACACCTGGACCCGA 1653
Db 582 AlaLysAlaAlaLeu-----SerAlaSerValAsnLeuValGlySerTyrThrThrAsp 600
QY 1654 AGTGAA---CAACGTGATGTGTATACAAAGGTGGCGCCGCTGAGTTATACCCCTGAACAC 1710
Db 601 AlaGluTyrThrThrAspThrThrTyrLysGlyAsnThrProAlaGlnValProLysHis 620
QY 1711 ATGTGTAATCGGAACCTGAAGTGCAGATCACCGAAGAGGTGGCATCTGCTG----- 1764
Db 621 MetAlaSerLeuTrpGlyAspTyrThrLeuPheAspGlyAlaLeuSerGlyLeuThrLeu 640
QY 1765 -----GGTCCCGTTATCCGCGGGAACACACACGCTTCCACCCAGATTAATTCGTCACTG 1818
Db 641 GlyThrGlyValArgTyrThrGlySerSer----- 650
QY 1819 AGCGCTGTACAGAAAGTGTATGATGATGAGAAGGAGATACCTGAAG-----GCCTGG 1872
Db 651 -----TyrGlyAspProAlaAsnSerPheLysValGlySerTyr 663
QY 1873 ACGGTGGTGATGACGCTGTGCTGGAGATG-----ACGGATGCCCTGACGCTGAAT 1926
Db 664 ThrValValAspAlaLeuValArgTyrAspLeuAlaArgValGlyMetAlaGlySerAsn 683
QY 1927 GCTGCG-----GTGTAATACCTGCTCAACAGGATTAC 1959
Db 684 ValAlaLeuHisValAsnAsnLeuPheAspArgGluTyr 696

RESULT 15
US-10-282-122A-69342
; Sequence 69342, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
```

```
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69342
; LENGTH: 755
; TYPE: PRT
; ORGANISM: Pseudomonas syringae
US-10-282-122A-69342

Alignment Scores:
Pred. No.: 4,39e-17 Length: 755
Score: 293.50 Matches: 173
Percent Similarity: 36.71% Conservative: 117
Best Local Similarity: 21.90% Mismatches: 271
Query Match: 17.86% Indels: 229
DB: 12 Gaps: 40

US-10-625-972-4 (1-2091) x US-10-282-122A-69342 (1-755)
QY 4 CGAATAACACACTCTGGCTTCCTAGTCACTCCCTCTCGGATTTTCAGCAGCAGCATTA 63
Db 31 ArgIleGluThrAlaProGluAlaValGluGlySerLeuGluAlaThrThrfile 50
QY 64 GCTGTGCAGAGAGTGTGATGATT----- 87
Db 51 SerGlyAlaSerArgLeuGluThrAlaThrGlyProValThrGlyTyrValAlaThrArg 70
QY 88 ---GTCCTCGCATCCGGCTATGAGAAAAGTGAACAGCAGCAGCGCCGCTGCTG 144
Db 71 GlyLeuSerAlaThrLysThrAspThrAlaLeuIleGluThrProGlnSerIleSerVal 90
QY 145 ATTAGCCAGGAGGAATTCAGCTCCAGCCAGTACCAGCATCTGGCGAGGCTCTGAGATCA 204
Db 91 ValThrLysAspGlnMetLysAlaGlnGlyAlaGluAsnLeuSerGlnMetLeuArgTyr 110
QY 205 GTAGAGGGTGTGGATGTTCAAGTGTACGGGTAAACCGGAGGGCTGAA---ATCAGC 261
Db 111 SerAlaAlaValValProGluThrArgGlySerThrAlaSerArgLeuAspMetLeuSer 130
QY 262 ATCCGAGGAATG---CCAGCCAGTTACACGCTGATCTGATTTGTTGTTGTCAGGGC 318
Db 131 IleArgGlyPheSerProAlaLeuTyr-----LeuAspGlyLeuArgMetPro 146
QY 319 GGAACGAGTACGTGACTCCCAACCGTTTCTGCCATGAATACGGGTTTCATGCCCCCT 378
Db 147 AspAsnArgAspAlaAlaPro-----GlnLysAsp 156
QY 379 CTGGCGCCCATTCAGCGTATTGAGGTATTCAGGGGGCGGATGTCACACTGTATGGTCT 438
Db 157 ValPheAspLeuGluArgValGluValLeuArgGlyProAlaSerValLeuTyrGlyGln 176
QY 439 GATGCGATGGCGGTGTGTGAATATCATACCAAGAAGATGCAGACAATGGCTCTCT 498
Db 177 AlaSerProSerGlyValValAsnMetValSerLysLeuProThrGluThrProPheHis 196
QY 499 TCCGTCAATGCGAGGCTGAATCTGCAGGAAACCAATGGGTAAACAGCAGCAGCTTT 558
Db 197 GluIle-----GlyLeuThrTyrGlyThrPheAsnLys-----LysArgThrThrPhe 212
```


QY	559	AATTCTCGGACAGTGGTGCCCTTTGTGGATGATCTGTC-----ACCTGCAGGTACGCCGT	615
Db	213	AspPhe-----GlyGlyProIleAspAspGlnGlyValTyrSerTyrArgLeuSerGly	230
QY	616	AGCACACAACAGCGTCAAGGGTTTCATCGTGCATCATCTGACGCGATACAGCAGGACCGGT	675
Db	231	-----LeuTyrAspAspAlaAspGlyGln	238
QY	676	ATTCTTATCCCACCGAGTCAAGAATTATAATCTTGGTCGACGCTTTGACTGAAAGCGC	735
Db	239	IleGluHis--ThrGluThrArgGlnSerLeuAlaSerAlaPheThrTrpArgpro	257
QY	736	TCGAGCAGGAGTGTGCTCTGGTTTCATATGCATACCACCCCGCAGCGTTATGATAACCGG	795
Db	258	AspGluAspThrSerLeuThrLeu-----LeuAlaAsnLeuGlnSerAspProLys	274
QY	796	GATGGGCAACTGGGAGTCTG-----ACGGGGGA	825
Db	275	GlyAlaSerTyrGlySerMetProAlaTyr-GlySerValVallysSerProThrGlyArg	294
QY	826	-----TATGACCGACCTGGCTATGAG-----CGAACA AAA	858
Db	295	HistLeuAspPheAspPheTyrAspGlyGluYsaSnPheGluYsSerAspArgGluTyr	314
QY	859	ATTTCACGCTGCCTATGATCATCTTTCACCTTCGAACTGGAATCGTATCTGTAAC--	915
Db	315	HisAlaValGlyTyr-----LeuPheGluHisLeuAsnAsp	327
QY	916	---TGGAACGAGACAGAAAAT-----AAAGTCGTGAGCTGTGACCAAGTGA	960
Db	328	ValTrpThrLeuArgGlnAsnAlaArgTyrLeuArgSerGluGlyValTyrArgSerIle	347
QY	961	CTGAAGCGCAGCAAATCGGGGCTTCGCGGCAGCG-----CGGAG	1002
Db	348	Tyr-----AsnGlyTrpGlyThrLeu---GlnProaspTyrArgThrSerGluArgAla	364
QY	1003	CTTAAGGAATCGAACCTTATCTGAATTCATTACTGTTACC-----	1044
Db	365	ThrIleAlaThrAspValAsnLeuAspSerTyr-ThrIleAspAsnGlnMetGlnAlaAsn	384
QY	1045	-----CCTCTGGAGAAATCTCATCTCGTTACGTTGCGGGCGGAGTTTCAGAGC	1092
Db	385	PheAspThrGlyProLeu-----GlnHisThrLeuLeuGluGlyAlaAspTyrGlnAsn	402
QY	1093	TCGTCC-----ATGAAGACGGAGTTGTCTTCCGACGACAGGTCAAACTTTC-----	1140
Db	403	ThrSerThrAspThrLysAlaGlyTyrGlyIleGlyProThrLeuAspIlePheAspPro	422
QY	1141	-----CGG	1143
Db	423	ValTyrGlySerProValGluValProAlaPheThrGluSerSerThrGlnArgAspGln	442
QY	1144	CAGAAAAAGCTGGTCGGTATTTCGTAGAGTAGTGGCATCTCAGCGGATGCACCTTGGCGTG	1203
Db	443	GlnLysGlyLeuTyrLeuGlnGluGlnLeuLysTrp-----AspLysTrpValLeu	459
QY	1204	ACTCGGGCAGCCGCTATGAAACATANGCAATTCGGGGACACTTCAGTCGCGGTGCA	1263
Db	460	LeuMetGlyGlyArgTyr-----	465
QY	1264	TATCTGCTGGGATGTCGAGATGCTCGAGCTGAAAGCGGTGTGACCCGGGATAT	1323
Db	466	-----AspTrpAlaGluSerSerAsnSerSerThrAsnLeuThrSerSerThr	481
QY	1324	ARG-----GCACCCAGATGGGGCAGGTACATAAA	1353
Db	482	LysThrArgSerSerAlaAspSerGluAlaPheThrGlyArgLeuGlyLeuValTyrLeu	501
QY	1354	GGGATTAGTGGTGTG-----TCCGGG	1374
Db	502	PheAspAsnGlyLeuAlaProTyrIleSerTyrSerGluSerPheGluProGlnSerGly	521
QY	1375	CAGGGAAAAACAAATCTACTTGGTATACCCCGACCTGAAGCCGGAAGACAGCGCTCAGTTAT	1434

Db	522	ThrGly----		-----	LeuGlyGlyAlaLeuPheAspProThrGluGlyThrGlnTyr	537
Qy	1435	GAGCGTGGGTGATTACGATAACCCCGCGTCTGAATGCCAAATGTCACAGGTTTATG		-----	LeuGlyGlyAlaLeuPheAspProThrGluGlyThrGlnTyr	1494
Db	538	GluLeuGlyIleLysTyr---GlnProProGlySerAsnSerPheIleThrAlaIle		-----	GluLeuGlyIleLysTyr---GlnProProGlySerAsnSerPheIleThrAlaIle	556
Qy	1495	ACTGATTC---TCCACCAAGATTGCTCTTTATTCATAAATGATAACACCAATAGC---		-----	ACTGATTC---TCCACCAAGATTGCTCTTTATTCATAAATGATAACACCAATAGC---	1548
Db	557	PheAspLeuArgSerAsnValLeuSerGlnAspProAsnGlyAsnLeuCysnGly		-----	PheAspLeuArgSerAsnValLeuSerGlnAspProAsnGlyAsnLeuCysnGly	576
Qy	1549	-----TATGTAAACAGCGGAAGCGCGTTCACGGGTGGAAATTTGCCGCCACA		-----	-----TATGTAAACAGCGGAAGCGCGTTCACGGGTGGAAATTTGCCGCCACA	1599
Db	577	AlaValCysGlnValcInThrGlyGluValGlnSerArgPheGluLeuGluGlyLys		-----	AlaValCysGlnValcInThrGlyGluValGlnSerArgPheGluLeuGluGlyLys	596
Qy	1600	TTGCCGCTGTGTCAGAG---GATGTCAGCTGTCACCTGAATTACCTGGACCCGGAAGT		-----	TTGCCGCTGTGTCAGAG---GATGTCAGCTGTCACCTGAATTACCTGGACCCGGAAGT	1656
Db	597	AlaSerLeuAsnAspAsnLeuAspIleThrAlaAlaTyrAlaTyrLeuAspAsnArgVal		-----	AlaSerLeuAsnAspAsnLeuAspIleThrAlaAlaTyrAlaTyrLeuAspAsnArgVal	616
Qy	1657	GAACAACGCTGAT-----		-----	GAACAACGCTGAT-----	1674
Db	617	ThrLysSerSerAsnValValAsnValThrSerGlyIleIleGlyAspAlaProGlyPro		-----	ThrLysSerSerAsnValValAsnValThrSerGlyIleIleGlyAspAlaProGlyPro	636
Qy	1675	-----AACAAAGTGGCGGTGAGTTATACCCCTGAACACATCGTGAATGCCGAA		-----	-----AACAAAGTGGCGGTGAGTTATACCCCTGAACACATCGTGAATGCCGAA	1725
Db	637	AspLeuThrThrLysGlyThrThrAlaProIleProArgHisThrAlaSerAlaTyr		-----	AspLeuThrThrLysGlyThrThrAlaProIleProArgHisThrAlaSerAlaTyr	656
Qy	1726	CTGAACCTGGCAGATCACCGAAGAGGTGCATCGCTG-----GGTGGCCGT		-----	CTGAACCTGGCAGATCACCGAAGAGGTGCATCGCTG-----GGTGGCCGT	1773
Db	657	ValAspTyrThrValHisAspGlyGlnLeuLysGlyAlaGlyValGlyGlyAlaArg		-----	ValAspTyrThrValHisAspGlyGlnLeuLysGlyAlaGlyValGlyGlyAlaArg	676
Qy	1774	TATCGCGGGAACACCCAGTTTACCACAGATATTTCGTCTACTGACGCGGTACAGNAG		-----	TATCGCGGGAACACCCAGTTTACCACAGATATTTCGTCTACTGACGCGGTACAGNAG	1833
Db	677	TyrValGlyAlaSer-----		-----	TyrValGlyAlaSer-----	681
Qy	1834	AAAGTGTATGATGAGAAAGGAAATACCTGAAA-----GCTGGACGGTGGTGATGCA		-----	AAAGTGTATGATGAGAAAGGAAATACCTGAAA-----GCTGGACGGTGGTGATGCA	1887
Db	682	-----TrpGlyAspGluAlaAsnThrLeuLysValProGlyTyrThrLeuPheAspAla		-----	-----TrpGlyAspGluAlaAsnThrLeuLysValProGlyTyrThrLeuPheAspAla	699
Qy	1888	GGTCTGTCTGGGAAGATG-----ACGGATGCCCTGACGCTCAATGCT		-----	GGTCTGTCTGGGAAGATG-----ACGGATGCCCTGACGCTCAATGCT	1929
Db	700	AlaValHisTyrAspIleProAsnIleSerAsnAlaMetAspAsnLeuArgLeuAlaLeu		-----	AlaValHisTyrAspIleProAsnIleSerAsnAlaMetAspAsnLeuArgLeuAlaLeu	719
Qy	1930	CGCGTGAATAACCTGCTCAACAAAGATTAC		-----	CGCGTGAATAACCTGCTCAACAAAGATTAC	1959
Db	720	AsnValThrAsnLeuAlaAsnLysGlnTyr		-----	AsnValThrAsnLeuAlaAsnLysGlnTyr	729

Search completed: October 13, 2004, 12:16:35
Job time : 178 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: October 13, 2004, 11:37:12 ; Search time 24.5 Seconds

(without alignments)
8888.057 Million cell updates/sec

Title: US-10-625-972-4

Perfect score: 3732
Sequence: 1 atgcgaataaccactctggc.....cgctgaactatcagttctga 2091

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlh
-Q/cn2_1/USPTC spool/US10625972/runat 12102004 130105 26806/app query.fasta 1.2247
-DB=SwissProt_42 -QPMT=fastan -SUFFIX=isp -MINMATCH=0.1 -LOOCPCL=0 -LOOCPXT=0
-UNITS=bits -START=1 -END=1 -MARIIX=blcosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10625972 @CN 1 1 29 @runat 12102004 130105 26806 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1052.5	28.2	652	IRGA_VIBCH	P27772 vibrio chol
2	832.5	22.3	663	CIRA_ECOLI	P17315 escherichia
3	625	16.7	746	FEPA_ECOLI	P05825 escherichia
4	592.5	15.9	746	PFEA_PSEAE	Q05098 pseudomonas
5	528.5	14.2	614	BTUB_ECOLI	P06129 escherichia
6	510.5	13.7	614	BTUB_SALTY	P37409 salmonella
7	373.5	10.0	700	YNCD_ECOLI	P76115 escherichia
8	355	9.5	687	HEMR_YEREN	P31499 versinia en
9	341	9.1	676	HMUR_YERPE	Q56989 versinia pe
10	339.5	9.1	673	FYUA_YERPE	P46359 versinia pe
11	332.5	8.9	673	FYUA_YEREN	P46360 versinia en
12	301.5	8.1	760	YBIL_ECOLI	P75780 escherichia
13	300.5	8.1	774	FECA_ECOLI	P13036 escherichia
14	291.5	7.8	723	V262_HAEIN	P43600 haemophilus
15	291	7.8	725	HXC2_HAEIN	P45357 haemophilus
16	282.5	7.6	720	FFTA_PSEAE	P42312 pseudomonas
17	280.5	7.5	702	FOXA_SALTY	Q56145 salmonella
18	278	7.4	747	FHUA_ECOLI	P06971 escherichia

19	266.5	7.1	710	1	FOXA_YEREN	Q01674 versinia en
20	262.5	7.0	743	1	BFRD_BORPE	P81549 bordetella
21	260	7.0	819	1	PUPA_PSEPU	P25184 pseudomonas
22	253	6.8	746	1	RHTA_RHIME	Q923Q5 rhizobium m
23	248.5	6.7	726	1	FATA_VIBAN	P11461 vibrio angu
24	244	6.5	810	1	HPUB_NEIMC	P96949 neisseria m
25	241.5	6.5	810	1	HPUB_NEIMA	Q91wa2 neisseria m
26	239	6.4	944	1	LBPA_NEIMA	Q91k44 neisseria m
27	233	6.2	729	1	FHUE_ECOLI	P14542 escherichia
28	231.5	6.2	732	1	IUTA_ECOLI	Q06987 neisseria g
29	230	6.2	908	1	TB12_NEIMB	Q01996 neisseria m
30	228.5	6.1	915	1	TB11_NEIGO	P31827 escherichia
31	225.5	6.0	790	1	YDDB_ECOLI	Q09056 neisseria m
32	225.5	6.0	911	1	TB11_NEIMB	Q09056 neisseria m
33	220	5.9	735	1	FCT_ERWCH	Q47162 erwinia chr
34	220	5.9	815	1	FPVA_PSEAE	P48632 pseudomonas
35	220	5.9	943	1	LBPA_NEIMB	Q06379 neisseria m
36	214	5.7	345	1	YEGA_HAEIN	O86241 haemophilus
37	213	5.7	1067	1	HGBB_HAEIN	Q9kiv1 haemophilus
38	209	5.6	826	1	PBUA_PSESP	Q08017 pseudomonas
39	207	5.5	687	1	VIUA_VIBCH	Q00964 vibrio chol
40	207	5.5	993	1	HGBC_HAEIN	Q9kiv0 haemophilus
41	205.5	5.5	1084	1	HGP3_HAEIN	P44836 haemophilus
42	198.5	5.3	809	1	PUPB_PSEPU	P38047 pseudomonas
43	196.5	5.3	999	1	HGP2_HAEIN	P44809 haemophilus
44	193.5	5.2	999	1	HGPB_HAEIN	O87296 haemophilus
45	193.5	5.2	1013	1	HGBA_HAEIN	Q9kiv2 haemophilus

ALIGNMENTS

RESULT 1
IRGA_VIBCH STANDARD; PRT; 652 AA.
ID IRGA_VIBCH
AC P27772; Q9KUP0;
DT 01-AUG-1992 (Rel. 23, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Iron-regulated outer membrane virulence protein precursor.
GN IRGA OR VC0475.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Classical Ogawa 395 / ATCC 39541 / Serotype O1;
RX MEDLINE=91023868; PubMed=1406279;
RA Goldberg M.B., Boyko S.A., Butters J.R., Stoebner J.A.,
Payne S.M., Calderwood S.B.;
RT "Characterization of a Vibrio cholerae virulence factor homologous to
the family of TonB-dependent proteins.";
RL Mol. Microbiol. 6:2407-2418(1992).
[2]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
McDonald L., Ueberlack T., Fleischmann R.D., Niernan W.C., White O.,
Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae".
RL Nature 406:477-483(2000).
[3]
RN
RP SEQUENCE OF 1-151 FROM N.A.
RC STRAIN=Classical Ogawa 395 / ATCC 39541 / Serotype O1;
RX MEDLINE=91072235; PubMed=2174861;
RA Goldberg M.B., Boyko S.A., Calderwood S.B.;
RT "Transcriptional regulation by iron of a Vibrio cholerae virulence

RT gene and homology of the gene to the *Escherichia coli* fur system.";

RL J. Bacteriol. 172:6863-6870 (1990).

CC -!- FUNCTION: INVOLVED IN THE INITIAL STEP OF IRON UPTAKE BY BINDING

CC FERRIC VIBRIOBACTIN, AN IRON CHELATING SIDEROPHORE THAT ALLOWS

CC V. CHOLERAE TO EXTRACT IRON FROM THE ENVIRONMENT.

CC -!- SUBCELLULAR LOCATION: Outer membrane.

CC -!- MISCELLANEOUS: REGULATION OF THE IRGA EXPRESSION IS NEGATIVELY

CC REGULATED AT THE TRANSCRIPTIONAL LEVEL BY IRON.

CC -!- SIMILARITY: Belongs to the tonB-dependent receptor family.

CC -----

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U72152; AAC4766.1; -

DR EMBL; AE004134; AAP93648.1; -

DR PIR; D82317; D82317.

DR PIR; S25265; S25265.

DR HSPP; P05825; IFEP.

DR TIGR; VC0475; -

DR InterPro; IPR000531; TonB boxC.

DR Pfam; PF00593; TonB dep Rec; 1.

DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.

DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.

DR Virulence; Outer membrane; iron transport; Transport; TonB box;

KW Signal; Receptor; Complete proteome.

FT SIGNAL 1 25 POTENTIAL.

FT CHAIN 26 652 IRON-REGULATED OUTER MEMBRANE VIRULENCE

FT SITE 33 40 TONB BOX.

FT SITE 635 652 TONB C-TERMINAL BOX.

FT CONFLICT 294 294 D -> G (IN REF. 1).

FT CONFLICT 448 448 K -> Q (IN REF. 1).

FT CONFLICT 502 502 A -> T (IN REF. 1).

SQ SEQUENCE 652 AA; 71669 MW; A95F82FEC072EC93 CRC64;

Alignment Scores:

Pred. No.: 3,916-68 Length: 652

Score: 1052.50 Matches: 258

Percent Similarity: 51.80% Conservative: 116

Best Local Similarity: 35.73% Mismatches: 223

Query Match: 28.20% Indels: 125

DB: 1 Gaps: 24

US-10-625-972-4 (1-2091) x IRGA_VIBCH (1-652)

QY 40 CTCGGATTTTCAGCCAGCAGCATAGCT-----GCTGCAGAGATGCTGATGATT 87

DB 17 LeuMetPheSerAlaSerAlaPheAlaGlnAspAlaThrLysThrAspGluThrMetVal 36

QY 88 GTCCTGGATCCGGCTATCAGAAAAGCTACTACGAGCGCCGAGTGTTCGTGATT 147

DB 37 ValThrAlaAlaGlyTyrAlaGlnValIleGlnAsnAlaProAlaSerIleSerValIle 56

QY 148 AGCCAGGAGGAATTCAGTCCAGCCAGTACCAGCATCTGGCGAGGCTCTGAGATCAGTA 207

DB 57 SerArgGluAspLeuSerArgTyrTyrArgAspValThrAspAlaLeuLysSerVal 76

QY 208 GAGGTGTGGATGTTCAAGTGGTACGGGTAAACCGGAGGCGGTGGAATCAGCATCGCA 267

DB 77 ProGlyValThrValThrGlyGlyAspThrThr-----AspIleSerIleArg 93

QY 268 GGAATGCCAGCAGTTACAGCTGATCTGATGATGTTGTTCTGTCGAGCGGAGCAGT 327

DB 94 GlyMetGlySerAsnTyrThrLeuIleLeuValAspGlyLysArgGln---ThrSerArg 112

QY 328 GACGTGACTCCCAACGGTTTCT---GCCATGAATACCGGTTTCATGCCCCCTCTGCC 384

DB 113 GlnThrArgProAsnSerAspGlyProGlyIleGlnGlnGlyTyrLeuProLeuGln 132

QY 385 GCCATTGAGCGTATTGAGGTTATCAGGGGGCCGATGTCACACATGTATGCTCTGATGCG 444

DB 133 AlaIleGluArgIleGluValIleArgGlyProMetSerThrLeuTyrGlySerAspAla 152

QY 445 ATGGCGGTGTGGTGAATATCATTACCAGAAAGATGCAGACAAATGGCTCTCTCCGTC 504

DB 153 IleGlyGlyValIleAsnIleIleThrArgIysAspGlnGlnTrpSerGlyAsnVal 172

QY 505 AATGACGGCTGAATCTGCAGGAAGCAACAAATGGGGTAACACACACCGCTTAATTC 564

DB 173 GlnLeuSerThrValValGlnGluAsnArgAlaSerGlyAspGluGlnSerAlaAspHe 192

QY 565 TGGAGCAGTGTCCCTTGTGGATGTTCTGCAGCTGCAGGTACGCGGTAGCACAA 624

DB 193 PheValThrGlyProLeu---SerAspAlaLeuSerLeuGlnValTyrGlyGlnThr 211

QY 625 CAGCGTCAG-----GGTTCATCGTCACTCACTGAGC 657

DB 212 GlnArgAspGluAspGluIleGluHisGlyTyrGlyAspLysSerLeuArgSerLeu 230

QY 658 GATACAGCAGCGCGCTATTCTTATCCACCGAGTCACAGAAATTAATCTTGGT--- 714

DB 231 -----ThrSerLysLeuAsnTyrGlnLeuAsnPro 240

QY 715 -----GCACGTCTTGACTGAAGCGCTCGAGCAGGATGTGCTCGTTGATG 765

DB 241 AspHisGlnLeuGlnLeuGluAlaGlyValSerAlaGlnAsp----- 254

QY 766 GATACACCGCGCAGCGCTTATGATAACCGGGATGGGCACTGGGAGTCTGACGGGGGA 825

DB 255 -----ArgGluAsnValGlyLysSerAlaGlnSerSerGlyCysArgGlyThr 271

QY 826 TATGACCGGACC---CTGCGCTATGAGCGAAACAAATTTACGCTGCTATGATCAT 882

DB 272 CysSerAsnThrAspAsnGlnTyrArgArgAsnHisValAlaValSerHisGln----- 289

QY 883 TTCACCTTCGGAACATCGAAA-----TCGTATCTGAACTGGAACGAG 924

DB 290 -----GlyAspTrpGlnAspValGlyLysSerAspThrTyrLeuGlnTyrGluGlu 306

QY 925 ACAGAAATAAAGTCTGCTGAGCTTGTACGCGAGTGTACTGAAGCGCAGCAATGGGGCTT 984

DB 307 AsnThrAsnLysSer----- 311

QY 985 GCGGTCAGCGCGGAGCTTAGGAATCGAACCTTATCTCTGAATTCATTCTGCTTACC 1044

DB 312 -----ArgGluMetSerIleAspAsnThrValPheLysSerThrLeuValAla 327

QY 1045 CCTCTGGAGAACTCTCATCTGTTACGGTGGGGGGGAGTTTCAGAGCTCGTCCATGAAA 1104

DB 328 ProIleGlyGlu---HisMetLeuSerPheGlyValGluGlyLysHisGluSerLeuGlu 346

QY 1105 GACGGAGTGTCTCTCCACGACAGCTGAAACT---TTCGGCGAGAAAAGCTGGTCGGTA 1161

DB 347 AspLysThrSerAsnLysIleSerSerArgTyrHisIleSerAsnThrGlnTrpAlaGly 366

QY 1162 TTTGCTGAGGATCAGTGGCATCTCAGGATGCACTTGGCTGCTGCTGCGGCGCGCTAT 1221

DB 367 PheIleGluAspGluTrpAlaLeuAlaGluGlnPheArgLeuThrPheGlyGlyArgLeu 386

QY 1222 GAACATCATGACAAATTCGGGGACATCTCAGTCCCGTGCATATCTGGTCTGGATG 1281

DB 387 AspHisAspLysAsnTyrGlySerHisPheSerProArgValTyrGlyValTrpAsnLeu 406

QY 1282 GCAGATGCTCGACGCTGAAGCGGTGTGACACGCGGATATAAGCCACCCAGTGGG 1341

DB 407 AspProLeuTrpThrValLysGlyValSerThrGlyPheArgAlaProGlnLeuArg 426

QY 1342 CAGCTACATAAAGGATTAGTGTGTCGGGCGAGGAAAAACAAATCTACTTGGTAAC 1401

DB 427 GluValThrProAspTrpGlyGlnValSerGlyGlyGly-----AsnIleTyrGlyAsn 444


```
QY 1402 CCCGACCTGAGCGGAGAGAGCGTCTAGTTATGAGCTGGGGTGTATTACGATAACCCC 1461
Db 1445 ProlapLeuLysProGluThrSerIleAsnLysGluLeuSerLeuMetTyrSerThrGly 464
QY 1462 CCCGCTGGAATGCAATGTCACAGGTTTATGACTGCTCTCCACAGAAATGCTCTCT 1521
Db 1465 SerGlyLeuAlaAlaSerLeuThrAlaPheHisAsnAspPheLysAspLysIleThrArg 484
QY 1522 TATTCATATAATGATAAC-----ACCAATAGCTAT----- 1551
Db 1485 ValAlaCysProAlaAsnIleCysThrAlaGlyProAsnGlnTyrGlyAlaAlaProThr 504
QY 1552 -----GTAACAGCGGAAAGCCCGGTTGCACCGTGTGGAATTTGCCGCGACATTCGCG 1605
Db 1505 TyrArgValAsnIleAspGluAlaGlnThrTyrGlyAlaGluAlaThrLeuSerLeuPro 524
QY 1606 CTGTGTCAGAGGATGTCACGTGTCACCTGAATACACCTGACCGGAGGAGCAACAGT 1665
Db 1525 Ile-----ThrGluSerValGluLeuSerSerTyrThrThrHisSerGluGlnLys 543
QY 1666 GATGCTGATACAAAAGTGGCGCGCTGAGTTATACCCCTGACACATGCTGAATCGGAAA 1725
Db 1544 SerGlyAsnPheAlaGlyArgProLeuLeuGlnLeuProLysHisLeuPheAsnAlaAsn 563
QY 1726 CTGAATGTCAGATCACCGAAGAGTGCATCATGCTGGTGGTCCGCTATCGCGGAAA 1785
Db 1564 LeuSerTrpGlnThrThrAspArgLeuAsnSerTrpAlaAsnLeuAsnTyrArgGlyLys 583
QY 1786 ACACACAGCTTCCACCCAGAAATATTCTGCTACGTAGCGGTGTACAGAGAAAGTGTATGAT 1845
Db 1584 -----GluMetGlnProGluGlyGlyAla 591
QY 1846 GAGAAAGGAGAAATACCTG---AAACCTGAGCGGTGGTGGATGTCAGGCTGTCTGCTGGAAG 1902
Db 1592 SerAsnAspPheIleAlaProSerTyrThrPheIleAspThrGlyValThrTyrAla 611
QY 1903 ATGACGAGTCCCTGACGCTGATGCTGGGGAATAACCTCTCAACAAGGATTACAGT 1962
Db 1612 LeuThrAspThrAlaThrIleLysAlaAlaValTyrAsnLeuPheAspGlnGluValAsn 631
QY 1963 GACGTGAGCCTGTACAGTCCCGTAAAGTACGCTGTATGCGCGGTATTACTTCACAGC 2022
Db 1632 -----TyrAla---GluTyr----- 635
QY 2023 GGATCATCAACACAGGATATGTATGATCACTGACGCGAATATGATGCTCGTGAACTAT 2082
Db 1636 -----GlyTyrValGluAspGlyArgTyrTrpLeuGlyLeuAspIle 650
QY 2083 CAGTTC 2088
Db 1651 Alaphe 652

RESULT 2
CIRA_ECOLI STANDARD; PRT; 663 AA.
AC P17315;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Colicin I receptor precursor.
GN Colicin OR CIR OR FEVA OR B2155.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89123100; PubMed=2644220;
RA Nau C.D., Konisky J.;
RT "Evolutionary relationship between the TonB-dependent outer membrane
RT transport proteins, nucleotide and amino acid sequences of the
RT Escherichia coli colicin I receptor gene."
RL J. Bacteriol. 171:1041-1047(1989).
```

```
RN [2]
RP REVISIONS.
RA Nau C.D., Konisky J.;
RL J. Bacteriol. 171:4530-4530(1989).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / BHE2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=9746617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1433-1474(1997).
[5]
RP SEQUENCE OF 1-59 FROM N.A., AND SEQUENCE OF 26-46.
RC STRAIN=BPR2;
RX MEDLINE=88058737; PubMed=3316180;
RA Griggs D.W., Tharp B.B., Konisky J.;
RT "Cloning and promoter identification of the iron-regulated cir gene
RT of Escherichia coli."
RL J. Bacteriol. 169:5343-5352(1987).
[6]
RP SEQUENCE OF 1-125 FROM N.A.
RC STRAIN=BPR2;
RX MEDLINE=92250419; PubMed=1315732;
RA Steffes C., Ellis J., Wu J., Rosen B.P.;
RT "The lysp gene encodes the lysine-specific permease."
RL J. Bacteriol. 174:3242-3249(1992).
[7]
RP SEQUENCE OF 1-21 FROM N.A.
RC MEDLINE=90264362; PubMed=2160948;
RA Griggs D.W., Kafka K., Nau C.D., Konisky J.;
RT "Activation of expression of the Escherichia coli cir gene by an
RT iron-independent regulatory mechanism involving cyclic AMP-cyclic
RT AMP receptor protein complex."
RL J. Bacteriol. 172:3529-3533(1990).
CC -!- FUNCTION: NOT YET KNOWN. POSTULATED TO PARTICIPATE IN IRON
CC -!- INDUCTION: BY IRON AND BY A CYCLIC AMP/CYCLIC AMP RECEPTOR
CC -!- PROTEIN COMPLEX.
CC -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```


FT SIGNAL 1 25
 FT CHAIN 26 663 COLICIN I RECEPTOR.
 FT SITE 31 38 TONB BOX.
 FT SITE 646 663 TONB C-TERMINAL BOX.
 FT CONFLICT 97 97 S -> D (IN REF. 6).
 FT CONFLICT 528 528 I -> N (IN REF. 1).
 SQ SEQUENCE 663 AA; 7395 MW; 2C68A45D4B5EE414 CRC64;

Alignment Scores:
 Pred. No.: 2,87e-52 Length: 663
 Score: 832.50 Matches: 234
 Percent Similarity: 47.45% Conservative: 120
 Best Local Similarity: 31.37% Mismatches: 237
 Query Match: 22.31% Indels: 155
 DB: 1 Gaps: 25

US-10-625-972-4 (1-2091) x CIRA_ECOLI (1-663)

QY 40 CTCGGATTTCAGCCAGCAGCATAGCTGCTGCA----- 72
 Db 10 ValGlyLeuCysLeuSerAlaIleSerCysAlaTrpProValLeuAlaValAspAsp 29
 QY 73 GAGGATGTGATGTTCTCCGATCCGGTATGAGAAAGCTGACTAACGCGCGCC 132
 Db 30 GlyGluThrMetValThrAlaSerSerValGluGlnAsnLeuLysAspAlaProAla 49
 QY 133 AGTGTCTCTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTACCAGCATCTGGCGGAG 192
 Db 50 SerIleSerValIleThrGlnGluAspLeuGlnArgLysProValGlnAsnLeuLysAsp 69
 QY 193 GCTCTGAGATCAGTAGAGGTGGGATGTT---GAAAGTGGTACGGGTAACACCGGAGG 249
 Db 70 ValLeuLysGluValProGlyValGlnLeuThrAsnGluGlyAspAsnArgLysGly--- 88
 QY 250 CTGGAATACAGATCCGAGGAATCCGAGCAGTACAGCTGATATGATGATGTT 309
 Db 89 -----ValSerIleArgGlyLeuAspSerSerTyrThrLeuIleValAspGlyLys 106
 QY 310 CGTCAGGCGGAGCAGTACGCTGACTCCCAACGGTTTTCTGCCATGATACCGGGTTC 369
 Db 107 ArgValAsnSerArgAsnAlaValPheArgHisAsnAspPheAspLeuAsn----- 123
 QY 370 ATGCCCTCTGCGCCCGATGAGCGTATTAGGTTATCAGGGGCGCGATGCCACACTG 429
 Db 124 TrpIleProValAspSerIleGluArgIleGluValValArgGlyProMetSerSerLeu 143
 QY 430 TATGGTCTGATCGATGGCGGTGGTGGTGAATATCATACCAGAAAGATCCAGACAA 489
 Db 144 TyrGlySerAspAlaLeuGlyGlyValValAsnIleIleThrLysLysIleGlyGlnLys 163
 QY 490 TGGCTCTCTCCGTCATGAGCGCTGAATCTGAGAAAGCAACAAATGGGGTAACAGC 549
 Db 164 TrpSerGlyThrValThrValAspThrThrIleGlnGluHisArgAspArgGlyAspThr 183
 QY 550 AGCCAGTTTAATTTCTGGAGAGTGTCCCTTTGGATGATCTGTACGCTGAGGTA 609
 Db 184 TyrAsnGlyGlnPhePheThrSerGlyProLeuIleAspGlyValLeuGlyMetLysAla 203
 QY 610 CGGGTAGCACACACAGCGT-----CAGGGTTTCATCGGTACATCATCTG 654
 Db 204 TyrGlySerLeuAlaLysArgGluLysAspAspProGlnAsnSerThrThrAspThr 223
 QY 655 AGGATACAGCAGCGCGGTATTCCTTATCCACGAGTCAAGAAATTATTAATCTTGGT 714
 Db 224 GlyGluThr----- 226
 QY 715 GCACGCTTACCTGGAAGCGTCGGAGCAG-----GATGCTCTGGTTTGATATG 765
 Db 227 ProArgIleGlyGlnPheSerArgAspGlyAsnValGluPheAlaTrp----- 243
 QY 766 GATACACCGCGGAGCGTTATGATACCGGGATGGCAACTCGGGGAGTCTGACGGGGGA 825
 Db 244 ---ThrProAsnGlnAsnHisAspPheThrAlaGlyTyr-----Gly 256

QY 826 TATGACCG-----ACCTGCGCTATGACGGAACAAA 858
 Db 257 PheAspArgGlnAspArgAspSerSerLeuAspLysAsnArgLeuGluArgGlnAsn 276
 QY 859 ATTTACAGCTGGCTATGATCATCTTTCCACCTTCGGAACATGGAATCGTATCTGAACCTG 918
 Db 277 TyrSerValSerHisAsnGlyArgTrpAspTyrGlyThrSerGluLeuLysTyrTyrGly 296
 QY 919 AACGAGACAGAAATAAAGGTCGTGAGCTTACGCGAGTGTACTGAAGCGCACAATAGG 978
 Db 297 GluLysValGluAsnLysAsn----- 303
 QY 979 GGGCTTCGCGTACGCCCGCGGAGCTT---AAGGAATCGAACCTTATCTCGAATTCATTA 1035
 Db 304 -----ProGlyAsnSerSerProIleThrSerGluSerAsnThrValAspGlyLysTyr 321
 QY 1036 CTGCTTACCCCTCTGGGAGATCTCATCTGGTACGTTGGGGGGGAGTTTCAGAGCTCG 1095
 Db 322 ThrLeuProLeuThrAlaIleAsnGlnPheLeuThrValGlyGlyGluTrpArgHisAsp 341
 QY 1096 TCCATGAAGACGGAGTTGCTTTCCGACGACAGTGAATTTCCGCGCAGAAAGC--- 1152
 Db 342 LysLeuSerAspAlaValAsnLeu-----ThrGlyGlyThrSerSerLysThrSerAla 359
 QY 1153 -----TGGTGGTATTGCTGAGGATGAGTGGCATCTCAGGATGACACTGCGCTGACT 1206
 Db 360 SerGlnTyrAlaLeuPheValGluAspLutrpArgIlePheGluProLeuAlaLeuThr 379
 QY 1207 CGCGGACGCGCTATGAACATCATGACAAATTCGGGGGACACTTCAGTCCGCGTCATAT 1266
 Db 380 ThrGlyValArgMetAspAspHisGluThrTyrGlyGluHisTrpSerProArgAlaTyr 399
 QY 1267 CTGCTCTGGATGGGAGATGCTGACGCTGAAAGCGGTGTGACCCGCGGATATAAG 1326
 Db 400 LeuValTyrAsnAlaThrAspThrValThrValLysGlyGlyTrpAlaThrAlaPheLys 419
 QY 1327 GCACCCAGATGGGCGAGCTACATAAGGATAGTGTGTGTCGGGCGAGGGAACAA 1386
 Db 420 AlaProSerLeuLeuGlnLeuSerProAspTrpThrSerAsnSerCysArgGlyAlaCys 439
 QY 1387 AATCTACTTGTTAAACCCGACCTGAAGCCGGAAGAGAGCGTCTAGTTATAGGCTGGGTG 1446
 Db 440 LysIleValGlySerProAspLeuLysProGluThrSerGluSerTrpGluLeuGlyLeu 459
 QY 1447 TATTACGATAACCC-----GCCGGTCTGAATGCCAATGTCACAGGTTTATG 1494
 Db 460 TyrTyrMetGlyGluGluGlyTrpLeuGluGlyValGluSerSerValThrValPheArg 479
 QY 1495 ACTGACTTCTCCACACAGATTGCTCTTATTCCATAAATGATTAACACCAAT----- 1545
 Db 480 AsnAspValLysAspArgIle---SerIleSerArgThrSerAspValAsnAlaAlaPro 498
 QY 1545 ----- 1545
 Db 499 GlyTyrGlnAsnPheValGlyPheGluThrGlyAlaAsnGlyArgArgIleProValPhe 518
 QY 1546 AGCTATCTAAACAGCGGAAAGCGCGGTTCGACCGTGTGGAATTTCCGCGCACATTGCCG 1605
 Db 519 SerTyrTyrAsnValAsnLysAlaArgIleGlnGlyValGluThrGluLeuLysIlePro 538
 QY 1606 CTGTGCTCAGAGATGTCAGCTGTCTCACTGAATTCACCTGGACCCGAGTGAACACCTG 1665
 Db 539 ---PheAsnAspGluTrpLysLeuSerIleAsnTyrThrTyr-----AsnAspGlyArg 555
 QY 1666 GAT-----GTTGATAACRAAGTGGCGCTGAGTTATACCCCTGAACACAGCTGGTG 1716
 Db 556 AspValSerAsnGlyGluAsnLys-----ProLeuSerAspLeuProPheHisThrAla 573
 QY 1717 AATCGGAACTGAAGTGGCAG-----ATCACCGAAGAGTGGCATCATGCTGGTGGTGCC 1770
 Db 574 AsnGlyThrLeuAspTrpLysProLeuAlaLeuGluAspTrpSerPheTyrValSerGly 593

corresponding to the 12.7-28.0 min region on the linkage map.";

RNA Res. 3:137-155(1996).

[5]

SEQUENCE OF 1-77 FROM N.A.

RP MEDLINE=89066678; PubMed=2974033;

RN Pettis G.S., Brickman T.J., McIntosh M.A.;

RA "transcriptional mapping and nucleotide sequence of the Escherichia

RT coli fepA-fes enterobactin region. Identification of a unique

RT iron-regulated bidirectional promoter.";

RL J. Biol. Chem. 263:18857-18863(1988).

RN [6]

MOLECULAR ANALYSIS.

RP MEDLINE=90354449; PubMed=2201687;

RN Armstrong S.K., Francis C.L., McIntosh M.A.;

RA "Molecular analysis of the Escherichia coli ferric enterobactin

RT receptor fepA.";

RL J. Biol. Chem. 265:14536-14543(1990).

RN [7]

X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).

RP MEDLINE=99101384; PubMed=9886293;

RN Buchanan S.K., Smith B.S., Venkatramani L., Xia D., Esser L.,

RA Palnitkar M., Chakraborty R., van der Helm D., Deisenhofer J.;

RA "Crystal structure of the outer membrane active transporter FepA from

RT Escherichia coli.";

RL Nat. Struct. Biol. 6:56-63(1999).

RN [8]

FUNCTION. THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON

CC UPTAKE BY BINDING FERRITEROBACTIN (FE-ENT), AN IRON CHELATIN

CC SIDEROPHORE THAT ALLOWS E. COLI TO EXTRACT IRON FROM THE

CC ENVIRONMENT. FEPA ALSO ACTS AS A RECEPTOR FOR COLICINS B AND D.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.

CC -!- SIMILARITY: Belongs to the tonB-dependent receptor family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR ENBL; M13748; AAA65994.1; -.

DR ENBL; A8000163; AAC73685.1; -.

DR ENBL; U82598; AAB40783.1; ALT_INIT.

DR ENBL; D90700; BAA35225.1; -.

DR ENBL; J04216; BAA23756.1; -.

DR PIR; F64791; QRECFQ.

DR PDB; 1FEP; 13-JAN-99.

DR SWISS-2DPAGE; P05825; COLI.

DR ECO2DEASE5; D079.0; 6TH EDITION.

DR EcoGene; EG10293; fepA.

DR InterPro; IPR000531; TonB_boxC.

DR Pfam; PF00593; TonB_dep_Rec; 1.

DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.

DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.

KW Transmembrane; Outer membrane; Iron transport; TonB box;

KW Signal; Receptor; 3D-structure; Complete proteome.

FT SIGNAL 1 22

FT CHAIN 23 746 FERRITEROBACTIN RECEPTOR.

FT SITE 34 41 TONB_BOX.

FT SITE 729 746 TONB_C-TERMINAL_BOX.

FT CONFLICT 152 152 A -> R (IN REF. 1).

FT CONFLICT 403 403 MISSING (IN REF. 1).

FT HELIX 40 44

FT TURN 45 46

FT TURN 48 49

FT STRAND 50 54

FT HELIX 55 60

FT TURN 67 67

FT HELIX 68 71

FT TURN 72 73

FT TURN 75 76

FT TURN 77 81

FT STRAND 77 84

FT TURN 85

FT STRAND	737	746	746 AA; 82107 MW; 09348AAB1C29710A CRC64;
FT STRAND	737	746	
SQ SEQUENCE	746 AA; 82107 MW; 09348AAB1C29710A CRC64;		
Alignment Scores:			
Pred. No.:	2,738-37	Length:	746
Score:	625.00	Matches:	299
Percent Similarity:	45.25%	Conservative:	134
Best Local Similarity:	27.57%	Mismatches:	285
Query Match:	16.75%	Indels:	130
DB:	1	Gaps:	29
US-10-625-972-4 (1-2091) x FEPA_ECOLI (1-746)			
QY 73	GAGATGTGATGATGCTCTCGGCATCCGGCTATGAGAAAAAGCTACTAACACACCGCCG 132		
DB 33	AspAspThrIleValValThrAlaAla-----GluGlnAsnLeu-----GlnAlaPro 48		
QY 133	AGTGTTCGTGATTAGCCAGGAGGAATTCAGTCCAGCCAGTAC---CACGATCTGGCG 189		
DB 49	GlyValSerThrIleThrAlaAspGluIleArgLysAsnProValAlaArgAspValSer 68		
QY 190	GAGGCTCTGAGTCAGTACAGGGGTGTGGATGT-----GAAAGTGTACGGGTAAAAACC 243		
DB 69	LysIleIleArgThrMetProGlyValAsnLeuThrGlyAsnSerThrSerGlyGlnArg 88		
QY 244	GGA---GGGCTGGAAATCAGCATCCGAGGAATCCAGCCAGTACACCTGATCTGATT 300		
DB 89	GlyAsnAsnArgGlnIleAspIleArgGlyMetGlyProGluAsnThrLeuIleLeuIle 108		
QY 301	GATGGT-----GTTCTGTCAGGGCGGAGGAGCAGTACGTGACT 336		
DB 109	AspGlyLysProValSerSerArgAsnSerValaGlnGlyTrpArgGlyGluArgAsp 128		
QY 337	CCCAACGGTTTTTCTGCCATGAATACCGGGTTTCATGCCCCCTCTGCGCCCATTCAGCGT 396		
DB 129	ThrArgGly-----AspThrSerTrpValProPro---GluMetIleGluArg-143		
QY 397	ATTGAGGTTATCAGGGGCGGATGTCACACTGTATGCTCTGATCGCATGGCGGTGTG 456		
DB 144	LleGluValIleuArgGlyProAlaAlaAlaArgTrpGlyAsnGlyAlaAlaGlyVal-163		
QY 457	GTGAATATCATTCACAGAAAGAAATGCAGACAAATGGCTCTCTCCGTCATCGAGGGCTG 516		
DB 164	ValAsnIleIleThrLysLysGlySerGlyGluTrpHisGlySerTrpAspAlaTrpPhe 183		
QY 517	AATCTGCAGAAAGCAACAAATGGGTAAACAGCCAGTCTTAATTTCTGGAGCAGTGGT 576		
DB 184	AsnAlaProGluHisLysGluGluGlyAlaThrLysArgThrAsnPheSerLeuThrGly 203		
QY 577	CCCTTGTGGATGATTCGTGACGCTGCGGTAGCCGTAGCAGACACACAGCGTCAGCGT 636		
DB 204	ProLeu---GlyAspGluPheSerPheArgLeuTrpGlyAsnLeuAspLysThrGlnAla 222		
QY 637	TCATCG-----GTCATCATCTGACCGGATACAGCAGCGACCGCGT 675		
DB 223	AspAlaTrpAspIleAsnGlnGlyHisGlnSerAlaArgAlaGlyThrTyrrAlaThrThr 242		
QY 676	ATTCTTTATCCACCGAGTCACAG--AATTATATCTTGGTCACCGCTTCTGACTGGAAG 732		
DB 243	LeuProAlaGlyArgGluGlyValIleAsnLysAspIleAsnGlyValValArgTrpAsp 262		
QY 733	GGGTCCGACGAGTGTCTGTGTTTGATATGGATACACCCCGGACG-----780		
DB 263	PheAlaProLeuGlnSerLeuGluLeuAlaGlyTyrrSerArgGlnGlyAsnLeuTyrr 282		
QY 781	-----CGTTATGATAACCGGATCGGCACCTGGGGATCTCGCGGGGGATAT 828		
DB 283	AlaGlyAspThrGlnAsnThrAsnSerAspSerTyrrThrArgSerLysTyrrGly----- 300		
QY 829	GACCGGACCTCGGC---TATGAGCGCAACAAAAATT-----TCAGTGGCTATGAT 876		
DB 301	AspGluThrAsnArgLeuTyrrArgGlnAsnTyrrAlaLeuThrTrpAsnGlyGlyTrpAsp 320		

DR Pfam: PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
 KW Complete proteome.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 746 FERRIC ENTEROBACTIN RECEPTOR.
 FT SITE 39 44 TONB BOX.
 FT SITE 729 746 TONB C-TERMINAL BOX.
 SQ SEQUENCE 746 AA; 80967 MW; ADD4FB0CB22C3251 CRC64;

Alignment Scores:
 Pred. No.: 6.02e-35 Length: 746
 Score: 592.50 Matches: 211
 Percent Similarity: 43.28% Conservative: 124
 Best local Similarity: 27.26% Mismatches: 284
 Query Match: 15.88% Indels: 155
 DB: 1 Gaps: 30

US-10-625-972-4 (1-2091) x PF0A_PSEAE (1-746)

QY	55	AGCAGCATAGCTGTCGAGAGGATGTGATGCTCGGCATCCGGCTATGAGAAAG	114
Db	32	SerValIleGluLeuGlyGluInThrValAlaThrAla	47
QY	115	CTGACTAACCGCAGCGCCAGTCTTCTGTGATTAGCCAGGAGAAATG	171
Db	48	GluThrLysGlnAlaProGlyValSerIleIleThrAlaGluAspIleAlaLysArgPro	67
QY	172	CAGTACACGATCGCGGAGGCTCTGAGATCAGTAGAGGTGTGATGTT	225
Db	68	ProSerAsnAspLeuSerGlnIleIleArgThrMetProGlyValAsnLeuThrGlyAsn	87
QY	226	AGTGATACGGTAAACCGCA--GGCGTGAATCAGATCCGAGGAATGCCAGCAT	282
Db	88	SerSerGlyGlnArgGlyAsnAsnArgGlnIleAspIleArgGlyMetGlyProGlu	107
QY	283	TACACGCTGATCTGATGATGATGTT	318
Db	108	AsnThrLeuIleLeuValAspGlyLysProValSerSerArgAsnSerValArgTyrGly	127
QY	319	GGAAGCAGTACGTGACTCCCAACGGTCTTCTGCCATGAATACCGGTTCTATGCCCT	378
Db	128	TyrArgGlyGluArgAspSerArgGly	142
QY	379	CTGCGCCGATTCAGCGTATTCAGGTTATCAGGGGCGCATGTCACACATGCTGCTCT	438
Db	143	AlaAspGlnValGluArgIleGluValIleArgGlyProAlaAlaAlaArgTyrGlyAsn	162
QY	439	GATGCATGGCGGTGTGTGATATCATACAGAAAGAAATGCAGAAATGGCTCTCT	498
Db	163	GlyAlaAlaGlyValValAsnIleIleThrLysGlnAlaGlyAlaGluThrHisGly	182
QY	499	TCCGTCAATGCAGGCTGAATCTGCAGGAAAGCAACAATGGGTAAACAGCAGCGATT	558
Db	183	AsnLeuSerValTyrSerAsnProGlnHisLysAlaGluGlyAlaSerGluArgMet	202
QY	559	AATTTCTGGAGCAGTGGTCCCTTGTGGATATCTGTACAGCTCGAGTACCGGTAGC	618
Db	203	SerPheGlyLeuAsnGlyProLeuThr--GluAsnLeuSerTyrArgValTyrGlyAsn	221
QY	619	ACACACAG--	654
Db	222	IleAlaLysThrAspSerAspSerTrpAspIleAsnAlaGlyHisGluSerAsnArgThr	241
QY	655	AGCGATACACGACGCGGTATCTTATCCACCGAGTCA--CAGAATATATATCTT	711
Db	242	GlyLysGlnAlaGlyThr--LeuProAlaGlyArgGluGlyValArgAsnLysAspIle	260
QY	712	GGTGACGCTCTGACTGGAAGCGTCGAGCAGGATGCTCTGCTTCATATGATGATACC	771
Db	261	AspGlyLeuLeuSerTrpArgLeuThrProGluGlnThrLeuGluPheGluAlaGlyPhe	280

QY	772	ACCGCGCAG--	807
Db	281	SerArgGlnGlyAsnIleTyrThrGlyAspThrGlnAsnThrAsnSerAsnTyrVal	300
QY	808	GGGAGTCTGACGGGGGATATGACCGGACCTCGCTATGAGCGAAACAAAATTCAGCT	867
Db	301	LysGlnMetLeuGly--HisGluThrAsnArgMetTyrArgGluThrTyrSerVal	318
QY	868	GGCTATGATCATCTTCACCTTCGGAACATCGTAATCGTATCTGAACGAGACA	927
Db	319	ThrHisArgGlyGluTrpAspPheGlySerSerLeuAlaTyrLeuGlnTyrGluLysThr	338
QY	928	GAAATATAAGTCGTGAGCTGTACCGAGTGTACTGAAGCGGCAAAATGGGGCTTGC	987
Db	339	ArgAsn--	348
QY	988	GGT--	999
Db	349	GlyGlyThrGluGlyIlePheAspProAsnAsnAlaGlyPheTyrThrAlaThrLeuArg	368
QY	1000	GAGCTTAAG--	1050
Db	369	AspLeuThrAlaHisGlyGluValAsnLeuProLeuHis	382
QY	1051	GGAGAATCTCATCTCTGTTACGGTGGGGGGAGTTTCAGAGCTCGTCCATGAAGAC--	1107
Db	383	GlyTyrGluGlnThrLeuThrLeuGlySerGluTyrThrGluGlnLysLeuAspAspPro	402
QY	1108	--	1125
Db	403	SerSerAsnThrGlnAsnThrGluGluGlySerIleProGlyLeuAlaGlyLysAsn	422
QY	1126	ACAGGTGAACCTTCCGCGCAGAAAAGCTGTCGTATTTCTGAGGATAGTGGCATCTC	1185
Db	423	ArgSerSerSerSerAlaArgIlePheSerLeuPheAlaGluAspAsnIleGluLeu	442
QY	1186	ACGATGCATCTGCGTACTGCGGGCAGCGCTATGAACATCATGACAAATTCGGGGGA	1245
Db	443	MetProGlyThrMetLeuThrProGlyLeuArgTyrAspHisHisAspIleValGlyAsp	462
QY	1246	CACITTCAGTCCGGTGCATATCTGGTCTGGATGTGGCAGATCGCTGGACGCTGAAGGC	1305
Db	463	AsnTrpSerProSerLeuAsnLeuSerHisAlaLeuThrGluArgValThrLeuLysAla	482
QY	1306	GGTGTGACCGCATATAAGCAGCACCAGAAATGGGCGAGCTACATAAAGGAT--	1359
Db	483	GlyIleAlaArgAlaTyrLysAlaProAsnLeuTyrGlnLeuAsnProAspTyrLeuLeu	502
QY	1360	--	1404
Db	503	TyrSerArgGlyGlnGlyCysTyrGlyGlnSerThrSerCysTyrLeuArgGlyAsnAsp	522
QY	1405	GACCTGAAGCGGAAAGAGCGTCAGTTATGAGCTGGGGTGTATTACGATACACCGGCC	1464
Db	523	GlyLeuLysAlaGluThrSerValAsnLysGluLeuGlyIleGluTyrSerHisAsp--	541
QY	1465	GGTCTGAATGCCAATGTCACAGGTTTATGACTGACTCTCCCAACAGATT--	1515
Db	542	GlyLeuValAlaGlyLeuThrTyrPheArgAsnAspTyrLysAsnLysIleGluSerGly	561
QY	1516	--	1554
Db	562	LeuSerProValAspHisAlaSerGlyLysGlyAspTyrAlaAsnAlaIleTyr	581
QY	1555	--	1605
Db	582	GlnTrpGluAsnValProLysAlaValAlaGluGlyLeuGluGlyThrLeuThrLeuPro	601
QY	1606	CTG--	1650
Db	602	LeuAlaAspGlyLeuLysTrpSerAsnAsnLeuThrTyrMetLeu--	616
QY	1651	CGAAGTGAACACAGCTGATGATGATAAAAGGTGGCGCTGAGTTATACCCCTGAACAC	1710